

UNIVERSITÉ DE SHERBROOKE
Faculté de génie
Département de génie chimique et de génie biotechnologique

ÉVALUATION DE LA PERFORMANCE DU qPCR POUR LE DÉNOMBREMENT DE BACTÉRIES DANS LE MILIEU PHARMACEUTIQUE

Mémoire de Maîtrise en Génie Chimique

Charles GAUDREAULT

Jury : Nathalie FAUCHEUX
Ryan GOSSELIN
Joël SIROIS (directeur académique)
Joanny SALVAS (superviseur industriel)

« [...] et ainsi nous rendre comme maîtres et possesseurs de la nature [...] principalement aussi pour la conservation de la santé, laquelle est sans doute le premier bien et le fondement de tous les autres biens de cette vie»

— René Descartes, *Discours de la méthode*, sixième partie

RÉSUMÉ

La finalité du projet de recherche est de réduire le délai existant entre l'échantillonnage microbiologique des matières pharmaceutiques et le temps d'obtention du résultat des analyses; c'est-à-dire réduire le temps de cycle des tests microbiologiques. Concrètement, le projet de recherche vise à éliminer le temps d'attente du test de dénombrement des bactéries; ce temps est de 5 jours dus à l'incubation des microorganismes. Le temps de cycle des analyses microbiologiques est le goulot d'étranglement de la relâche de quarantaine des équipements et des matières premières. De plus, la découverte de contaminations se fait avec un temps de retard sur l'apparition de la contamination ce qui entraîne des pertes de produits et au final des coûts pour une usine qui relâche ses équipements de la quarantaine avant l'obtention du résultat. L'objectif du projet de recherche est d'évaluer la performance d'une méthode qPCR pour réaliser le dénombrement de bactéries dans le milieu pharmaceutique. La démarche consiste à évaluer la technologie selon les critères des instances réglementaires pharmaceutiques (limite de détection, précision, linéarité, robustesse, spécificité, etc.) et selon le principe d'équivalence avec la méthode traditionnelle. Les présents travaux de recherche permettent d'évaluer le potentiel de la méthode à être validée pour l'implantation dans une usine pharmaceutique. Les paramètres suivants ont aussi été évalués pour connaître leur influence sur la quantification et/ou leur relation avec la méthode traditionnelle : effet de l'état de croissance des cellules, effet de l'algorithme de quantification et effet des bactéries totales indépendamment de leur état de viabilité.

Le projet de recherche a démontré que la méthode de q-PCR avait le potentiel d'être validée en utilisant les critères pharmaceutiques en soulevant certains obstacles qui risquent de rendre la validation difficile. Le principal obstacle est le rapport variable entre le résultat de quantification par qPCR et celui de la méthode traditionnel. Cet obstacle peut être surmonté à condition d'avoir préalablement validé une méthode de préparation de cellules qui permet d'obtenir un rapport stable entre les unités de quantification de la méthode traditionnelle et de la méthode de qPCR. Il a été démontré que la préparation de cellules et son effet sur l'état de croissance de celles-ci sont à l'origine de ces variations. La gamme de produits apte à être traitée par la méthode de qPCR comprend les tests d'eau purifiée et les tests de liquides filtrables. La gamme de matières pouvant être analysée par la méthode de qPCR pourrait potentiellement être élargie à condition d'élargir nos connaissances sur l'effet des bactéries mortes et des bactéries viables non cultivables sur le contrôle microbiologique pharmaceutique. De plus, la gamme de matière traitée par la méthode pourrait être élargie avec le développement d'une méthodologie spécifique à la préparation de ces matières. L'étude a aussi démontrée que l'algorithme utilisé pour la quantification avait un effet important sur la précision, l'erreur, le biais et la robustesse. Un algorithme qui augmente spécifiquement la robustesse, de la méthode a été développé. Les conclusions de l'étude ne sont pas spécifiques au domaine pharmaceutique et peuvent s'étendre à d'autres domaines concernés par l'analyse microbiologique tel que l'alimentaire et les biotechnologies.

Mots clés : qPCR, PCR quantitatif, dénombrement, pharmaceutique, microbiologie rapide, bactéries, USP 1223, sonde universelle

REMERCIEMENTS

L'auteur aimerait prendre le temps de remercier ses collègues de Pfizer Montréal. L'auteur remercie premièrement sa superviseure, Joanny Salvas, qui a offert du temps, des conseils et du support tout au long de ce projet. Jean-Sébastien Simard qui a permis que ce projet soit possible. Ses collègues étudiants comme Philip Quinn, Emmanuel Vachon-Lachance et plusieurs autres ainsi que ses collègues de PASG Montréal avec qui les échanges d'idées et de service ont facilité l'avancement de ce projet. Le projet n'aurait pas pu être possible sans la collaboration, l'aide et les précieux conseils des analystes du laboratoire de microbiologies et de leurs superviseurs.

L'auteur remercie également ses professeurs du département de génie chimique et biotechnologique de l'Université de Sherbrooke. Notamment son directeur de maîtrise Joël Sirois pour le support académique du projet ainsi que Ryan Gosselin pour son aide concernant l'analyse de données multivariées, le traitement spectral et l'utilisation de MATLAB.

La collaboration et les judicieux conseils des scientifiques travaillant chez notre collaborateur industriel méritent d'être soulignés. Cependant, pour des raisons de confidentialité, leurs noms ainsi que le nom de l'entreprise ne seront pas mentionnés.

Finalement, l'aide financière du CRSNG, du FQRNT et de Pfizer par le biais de la bourse BMP-Innovation a rendu possible la réalisation de ce projet avec une rémunération raisonnable de l'auteur.

TABLE DES MATIÈRES

RÉSUMÉ	iii
REMERCIEMENTS	iv
LISTE DES FIGURES	viii
LISTE DES TABLEAUX	ix
LISTE DES ABRÉVIATIONS	x
1. INTRODUCTION	1
1.1. Mise en contexte et problématique	1
1.2. Définition du projet de recherche	2
1.3. Objectif du projet de recherche	3
1.3.1. Objectifs techniques	3
1.3.2. Objectifs scientifiques	3
1.3.3. Objectifs économiques	3
1.4. Contribution originale	4
2. qPCR — CADRE DE RÉFÉRENCE	5
2.1. Principes de base	5
2.1.1. Le PCR	5
2.1.2. Le PCR quantitatif (qPCR)	6
2.1.3. Variantes du PCR	6
2.2. Fluorescence et mélange réactionnel	7
2.2.1. Stratégie de fluorescence	7
2.2.2. Éléments du mélange réactionnel	8
2.3. Analyse de données	8
2.3.1. Détermination du cycle de quantification (C_q)	9
2.3.2. Modélisation de la courbe sigmoïde	11
2.3.3. Détermination de l'efficacité d'amplification (E)	12
2.3.4. Courbe standard et calcul de la quantité d'ADN	12
2.3.5. Analyse statistique	13
2.4. Contrôle de la qualité	14
2.4.1. Analyse des courbes de PCR	14
2.4.2. Analyse de la courbe standard	15
2.4.3. Analyse des contrôles	16
2.4.4. Analyse des répliques	16
3. DÉCOMPTE PAR QPCR — ÉTAT DE L'ART	17
3.1. Utilisation d'une sonde universelle pour les bactéries	17
3.2. Équivalence entre le qPCR et le décompte sur pétri	18
3.3. Limite de détection et limite de quantification	20
4. TECHNOLOGIE ET MÉTHODE À L'ESSAI	21
4.1. Description de la méthode	21
4.1.1. Vue d'ensemble de la méthode	21
4.1.2. Particularité de la méthode	22
4.1.3. Unité de quantification : UGC versus UFC	23
4.2. Procédure	24
4.2.1. Préparation des cellules	24
4.2.2. Procédure détaillée	25

5.	PERFORMANCES ANALYTIQUES	27
5.1.	Mise en contexte	27
5.2.	Méthodologie	28
5.2.1.	Exactitude — taux de récupération	28
5.2.2.	Spécificité — universalité	28
5.2.3.	Précision	29
5.2.4.	Linéarité.....	29
5.2.5.	Plage et limites de quantifications	30
5.2.6.	Robustesse interne	30
5.2.7.	Sommaire des essais	33
5.3.	Résultats	34
5.3.1.	Exactitude — taux de récupération	34
5.3.2.	Spécificité — universalité	35
5.3.3.	Précision	35
5.3.4.	Linéarité.....	37
5.3.5.	Plages et limites de quantification	38
5.3.6.	Robustesse interne	39
5.4.	Discussion	43
5.5.	Sommaire	44
6.	EFFET DE L'ÉTAT MÉTABOLIQUE	45
6.1.	Mise en contexte	45
6.2.	Méthodologie	45
6.2.1.	Terminologie	45
6.2.2.	Essai préliminaire	46
6.2.3.	Plan factoriel.....	46
6.3.	Résultats	47
6.3.1.	Essai préliminaire	47
6.3.2.	Plan factoriel.....	48
6.4.	Discussion	49
6.4.1.	Comportements observés.....	49
6.4.2.	Hypothèses retenues	50
6.4.3.	Autres hypothèses.....	52
6.5.	Sommaire	53
7.	EFFET DE L'ALGORITHME	54
7.1.	Mise en contexte	54
7.2.	Méthodologie	54
7.2.1.	Développement et évaluation comparative de l'algorithme.....	54
7.2.2.	Évaluation de l'impact sur la précision de la méthode de qPCR	55
7.3.	Résultats.....	56
7.3.1.	Développement et évaluation comparative de l'algorithme.....	56
7.4.	Principales conclusions de l'article	85
7.5.	Impact sur la précision de la méthode de qPCR	86
7.5.1.	Résultats	86
7.5.2.	Analyse	87
7.6.	Sommaire	88

8.	ADAPTABILITÉ DE LA MÉTHODE	89
8.1.	Mise en contexte	89
8.2.	Méthodologie	89
8.2.1.	Poudres et comprimés	89
8.2.2.	Adaptation pour matières non filtrable	89
8.2.3.	Élimination de l'ADN extracellulaire	90
8.2.4.	Détermination de la quantité de matière utilisée	91
8.2.5.	Matières à l'essai	92
8.3.	Résultats	93
8.3.1.	Matières solides	93
8.3.2.	Matières liquides	96
8.4.	Discussion	97
8.4.1.	Matières non adaptées et pistes de solutions	97
8.4.2.	Stratégie pour l'analyse de produits finis	98
8.5.	Sommaire	98
9.	INFLUENCE DES BACTÉRIES TOTALES	100
9.1.	Mise en contexte	100
9.2.	Méthodologie	100
9.2.1.	Essais sur l'eau purifiée	100
9.2.2.	Essais sur les matières premières et produits finis	101
9.2.3.	Essais sur différents fournisseurs	101
9.3.	Résultats	102
9.3.1.	Essais sur l'eau purifiée	102
9.3.2.	Essais sur les matières premières et produits finis	103
9.3.3.	Essais sur différents fournisseurs	103
9.4.	Discussion	104
9.4.1.	Analyse des matières à haute teneur en UCG	104
9.4.2.	Analyse des matières à basse teneur en UCG	106
9.4.3.	Analyse de différents fournisseurs	106
9.5.	Sommaire	107
10.	CONCLUSION	108
10.1.	Sommaire des travaux	108
10.2.	Retour sur les objectifs et la question de recherche	111
10.2.1.	Objectifs techniques	111
10.2.2.	Objectifs scientifiques	111
10.2.3.	Objectifs économiques	111
10.2.4.	Question de recherche	112
10.3.	Contributions originales	113
10.4.	Nouvelles perspectives de recherches	113
	ANNEXE A	114
	ANNEXE B	115
11.	RÉFÉRENCES	116

LISTE DES FIGURES

Figure 2.1. Étape d'un cycle de PCR.....	6
Figure 2.2. Courbes typiques d'un PCR.....	10
Figure 2.3. La courbe de fluorescence décrite par quatre paramètres	11
Figure 2.4. Division de la courbe de PCR	14
Figure 5.1. ETR pour la plage inférieure de concentration pour les essais de 2012	36
Figure 5.2. ETR pour les 17 suspensions réalisées avec le dernier protocole en vigueur	36
Figure 5.3. Tests de linéarité.	37
Figure 5.4. Effet de la concentration sur le taux de récupération.	39
Figure 5.5. Résultats de l'analyse uni-variée.	40
Figure 5.6. Moyenne des niveaux des facteurs étudiés pour l'analyse de la robustesse	40
Figure 5.7. Effet du type nettoyage d'une pièce réutilisable sur le résultat des contrôles.	42
Figure 6.1. Interaction entre l'âge de la culture sur pente et l'âge de la suspension	48
Figure 6.2. Effet hypothétique de l'âge de la culture et de la suspension	51
Figure 6.3. Effet de l'ordre des échantillons sur le résultat final	52
Figure 8.1. Arbre décisionnel pour le traitement des troubles de colmatage du filtre.....	90
Figure 8.2. Effet combiné du CaCO_3 et de la maltodextrine	94
Figure 8.3. Récupération avec lavage acide	95

LISTE DES TABLEAUX

Tableau 2.1. Relation C_q , [ADN] et ETR.....	13
Tableau 3.1. Équivalence entre le qPCR et la méthode conventionnelle dans la littérature	19
Tableau 5.1. Traitement de l'essai uni-varié pour évaluer la robustesse	31
Tableau 5.2. Plan fractionnaire 2^4 -1 en duplicata pour évaluation de la robustesse	32
Tableau 5.3. Sommaire des essais de l'évaluation analytique	34
Tableau 5.4. Taux de récupération par la méthode qPCR.....	35
Tableau 5.5. Valeur de la régression linéaire	37
Tableau 5.6. Plan factoriel 2×2^3 — ANOVA.....	41
Tableau 6.1. Plan 2×3^2 : traitements	47
Tableau 6.2. Influence de l'âge de la suspension bactérienne	47
Tableau 6.3. Plan factoriel 2×2^3 16 essais — ANOVA.....	49
Tableau 7.1. Effet des différents algorithmes sur les tests au niveau de la précision.....	87
Tableau 8.1. Matière à l'essai pour l'adaptation de la méthode	92
Tableau 8.2. Adaptation du protocole pour les matières solides.....	93
Tableau 8.3. Adaptation du protocole pour les matières liquide.....	96
Tableau 9.1. Sommaire des tests d'eau	102

LISTE DES ABRÉVIATIONS

4PLM : Cycle de quantification basé sur modélisation de la courbe sigmoïde utilisant 4 paramètres
 ADN : Acide désoxyribonucléique
 ADN_c : Brin d'ADN complémentaire
 ADN_e : ADN extracellulaire
 ARN : Acide ribonucléique
 ATP : Adénosine triphosphate
 CCD: *Coupled-charged-device*
 C_q : Cycle de quantification
 Ct : *Cycle Threshold*; cycle de quantification basé sur la méthode du threshold
 CV : Coefficient de variation; valeur absolue de l'écart-type divisé par la moyenne
 C_{y0} : Cycle de quantification basé sur la méthode de la tangente qui croise l'abscisse à l'origine (Guescini, 2008)
 DoE : *Design of Experiment*, ou plan d'expérience en français; concept statistique
 E : Efficacité d'amplification; valeur calculée entre les cycles de PCR ou moyennée à l'aide d'une courbe standard
 EP : *European Pharmacopea*
 ET : Écart-Type, en anglais SD pour *Standard Deviation*
 ETR : Écart-type relatif
 G_c : *Gene copies*, copies de gène en français
 LED : *Light-emitting diode*; diode émettant de la lumière en français
 LGSG : Lissage et Différentiation de Savitzky-Golay
 LOD : *Limit of detection* ; limite de détection en français
 LOQ : *Limit of quantification*; limite de quantification en français
 LQ_I : Limite inférieure de quantification
 LQ_s : Limite supérieure de quantification
 MPN : *Most probable number* ; nombre le plus probable en français
 MS : *Mass spectrometry*; spectrométrie de masse en français
 PA: Pente âgé; culture bactérienne sur pente d'agarose qui à été vieillis délibérément
 PASG : *Process analytical science group*
 PAT : *Process analytical technology*; technologies d'analyses de procédés en français
 PCR : de l'anglais *polymerase chain reaction*
 PF : Pente fraîche; culture bactérienne sur pente fraîchement inoculée
 PV : Pente vierge; pente d'agarose non inoculée
 qPCR : *Quantitative PCR*, aussi appelé *real time PCR* (RT-PCR) dont l'acronyme porte à confusion avec le *reverse transcriptase PCR* (RT-PCR)
 RB : *Relative bias*; biais relatif (BR) en français
 RE : *Relative error*; erreur relative (ER) en français
 RSD : *Relative standard deviation*; écart-type relatif (ETR) en français
 RT-PCR : *Reverse transcriptase PCR* (RT-PCR)
 RT-qPCR : *Reverse transcriptase quantitative PCR* (RT-qPCR)
 SA: Suspension âgé; suspension bactérienne dans de la saline qui à été vieillis délibérément
 SF : Suspension fraîche; suspension bactérienne dans de la saline fraîchement préparée
 SGSD : *Savitzky-Golay Smoothing and Differentiation*

UCG : Unité de copies de gènes

UFC : Unité formant colonie, en anglais *colony forming unit* (CFU)

USP: *United States Pharmacopeia*

Valeur-p : Probabilité d'obtenir la même valeur ou le même effet; concept statistique; *p-value* en anglais

TAC : *Total aerobic count*; décompte total des bactéries aérobies en français

TCYM : *Total count yeast and mold*; décompte total des levures et moisissures en français

1. INTRODUCTION

1.1. Mise en contexte et problématique

Dans le domaine pharmaceutique, le contrôle microbiologique des produits, matières premières, eau, surface et air est majoritairement effectué avec des méthodes de microbiologie conventionnelle. Dans le cas des usines fabriquant des produits non stériles, les tests peuvent prendre une semaine avant d’être complétés. Ce délai fait en sorte que les produits ne peuvent sortir de l’usine avant que le résultat ne soit connu. Dans le cas des matières premières, ils ne peuvent être utilisés avant l’obtention du résultat des tests. Ceci fait en sorte que l’entrepôt est surdimensionné pour avoir l’espace nécessaire pour mettre en quarantaine produits et matières premières. De plus, de tous les tests effectués sur les matières, les tests microbiologiques sont en général les plus longs.

La méthode traditionnelle consiste à prendre un échantillon contenant potentiellement des microorganismes et à l’ensemencer sur une gélose pour y compter les colonies. Le résultat est obtenu cinq jours plus tard pour le décompte du total des bactéries aérobies (*total aerobic count*, TAC) et sept jours pour le décompte du total des levures et moisissures (*total count yeast and mold*, TCYM) tel que précisé dans USP<61> [United States Pharmacopeia, 2012]. Le délai pour l’obtention des résultats est dû au temps d’incubation des microorganisme. Cependant, il existe des alternatives à la microbiologie traditionnelle. Plusieurs technologies existantes sur le marché ou en développement permettent de faire de la «microbiologie rapide». Ces technologies englobent entre autres l’ATP-bioluminescence, la cytométrie de flux, la cytométrie en phase solide et le PCR quantitatif (qPCR).

Le qPCR a le potentiel de remplacer les tests microbiologiques conventionnels pour les tests TAC. L’industrie pharmaceutique s’intéresse à ce genre d’application pour remplacer graduellement les boîtes de pétri par des tests plus rapides afin de rendre ses usines plus rapides et flexibles. Ces travaux auront pour but d’évaluer le potentiel du PCR quantitatif à remplacer les tests TAC pour les matières premières et produits finis du domaine pharmaceutique. La technologie devra être évaluée techniquement et économiquement. De

plus, le lien entre les bactéries cultivables et les bactéries dont l'intégrité cellulaire subsiste sans être cultivable sera étudié avec cet instrument.

Le projet de recherche s'insère dans les travaux du groupe PASG (Process Analytical Science Group) de Pfizer Montréal qui travaille sur les PAT (Process Analytical Technologies). Le groupe s'intéresse aussi aux technologies de microbiologie rapide (*rapid microbiology method*, RMM) qui ne sont pas toutes des PAT par définition. Dans le cas idéal, il serait souhaitable que tous les tests microbiologiques soient remplacés par des PAT. Cependant la technologie actuelle ne le permet que pour quelques applications. Remplacer les tests TAC actuels par des technologies RMM serait un pas dans cette direction.

L'aboutissement du projet de recherche permettra à l'usine de Montréal de réduire le temps de cycle des analyses microbiologiques pour les produits et matières premières compatibles avec la méthode. D'un point de vue plus général, la conclusion tirée pour l'usine de Pfizer Montréal pourrait être adaptée à tout type d'endroit où l'on doit faire des analyses microbiologiques : l'industrie alimentaire, biotechnologique ou encore le domaine clinique.

1.2. Définition du projet de recherche

Tel que mentionné dans l'introduction, le temps d'attente est de cinq jours pour l'obtention des résultats avec la méthode actuelle de dénombrement du total des bactéries aérobies, ce qui est long. Ce délai ralentit la relâche des produits et de matières premières. L'hypothèse de recherche est qu'il y aurait un intérêt à remplacer la méthode actuelle par une méthode plus rapide afin de diminuer le temps de relâche des produits et des matières premières.

L'objectif de la maîtrise est d'évaluer une méthode de quantification de bactéries dans les échantillons pharmaceutiques basée sur le PCR quantitatif. Le projet de recherche doit répondre à la question suivante : la méthode de qPCR proposée peut-elle être utilisée pour remplacer les tests microbiologiques TAC en respectant les critères techniques et économiques d'une usine pharmaceutique telle que celle de Pfizer Montréal ?

1.3. Objectif du projet de recherche

1.3.1. Objectifs techniques

- I.** Évaluer si le qPCR peut être utilisé pour remplacer les tests TAC
 - Évaluer la méthode au niveau des normes pharmaceutiques USP* et EP**
- II.** Déterminer pour quel type de matière le qPCR peut remplacer les tests TAC
 - Adapter la méthode pour l'analyse de différentes matières
 - Évaluer le principe d'équivalence comparativement à la méthode traditionnelle

1.3.2. Objectifs scientifiques

- I.** Évaluer l'impact de la viabilité des cellules sur la méthode
 - Évaluer le biais créé par la présence de bactéries mortes ou non cultivables dans les matières pharmaceutiques
 - Évaluer l'impact de l'état de croissance des cellules sur le biais entre les résultats sur pétri et les résultats sur qPCR
- II.** Évaluer l'impact de la méthode de quantification sur le résultat de quantification

1.3.3. Objectifs économiques

- I.** Déterminer la rentabilité du projet
 - Évaluer le coût des tests pour l'utilisation de la nouvelle méthode
 - Se servir de l'usine de Montréal comme cas de base et évaluer si l'implantation de l'appareil et de la méthode est justifiée économiquement

* USP : United States Pharmacopeia

** EP : European Pharmacopeia

1.4. Contribution originale

1. Les travaux présentés dans ce mémoire ont permis de faire l'évaluation d'une méthode de qPCR ayant pour objectif de remplacer le test du compte des bactéries totales pour le domaine pharmaceutique. L'évaluation est basée sur les lignes directrices de la « United States Pharmacopeia » en ce qui a trait à l'usage de méthodes microbiologiques alternatives.
2. Les travaux ont aussi permis de soulever les lacunes en ce qui a trait à l'utilisation de la méthode traditionnelle comme base de comparaison. En effet, le dénombrement des unités formant une colonie est biaisée par le phénomène d'agrégation. Ceci rend difficile la comparaison de la méthode traditionnelle à des méthodes capables de compter chaque cellule individuellement indifféremment de l'agrégation.
3. Les travaux présentés ont aussi permis de développer un algorithme de quantification d'ADN dont le principal avantage est la robustesse face aux variations inter-essais et aux phénomènes d'inhibition.

2. qPCR — CADRE DE RÉFÉRENCE

2.1. Principes de base

Le PCR (polymerase-chain-reaction) est une technique de biologie moléculaire développée en 1983 par Kary Mullis [Mullis, 1993; Dorak, 2006]. Cette technique permet d'amplifier de façon géométrique une séquence d'ADN d'intérêt dans un échantillon. C'est au début des années 1990 que débutât le développement de la quantification à partir de réaction de PCR [Higuchi *et al.*, 1992; Higuchi *et al.*, 1993]. En 1996, Applied Biosystems inc. lança la première plateforme commerciale de PCR quantitatif (qPCR), ce qui fut une avancée dans le domaine de la biologie moléculaire, démocratisant ainsi la quantification de séquence d'ADN d'intérêt.

2.1.1. Le PCR

La technique est basée sur une réaction en chaîne qui duplique la séquence d'ADN d'intérêt jusqu'à épuisement du mélange réactionnel. La réaction est cyclique et à chaque cycle le nombre de brins d'ADN est doublé. Chaque cycle débute avec l'ADN double brin qui subira trois étapes de transformation intermédiaires passant par l'ADN simple brin pour terminer le cycle avec l'ADN double brin. La première étape est une étape de dénaturation thermique à environ 95 °C dans le but de briser les ponts hydrogène liant les deux brins d'ADN. La seconde étape est celle d'hybridation de l'amorce à la séquence d'ADN; l'amorce s'hybride lorsque l'on refroidit le mélange à une température légèrement inférieure à la température de fusion de l'amorce (entre 50 et 65 °C). Cette amorce permettra à l'ADN polymérase de se fixer en aval de celle-ci sur le brin d'ADN afin de synthétiser l'ADN complémentaire à celui-ci. Ensuite, la température est élevée à l'optimum de la polymérase autour de 75 °C, dépendamment du type d'ADN polymérase, et l'ADN est synthétisé. À la fin d'un cycle, l'ADN aura été doublé. Les étapes sont représentées dans la figure 2.1.

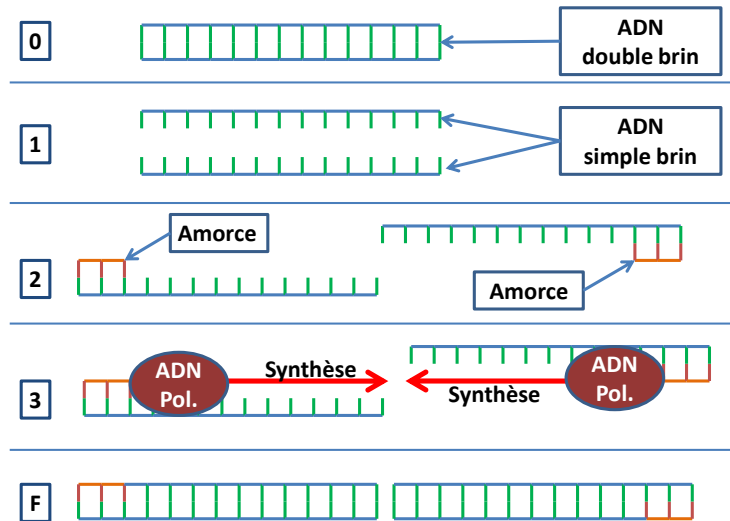


Figure 2.1. Étape d'un cycle de PCR

2.1.2. Le PCR quantitatif (qPCR)

Le PCR quantitatif (qPCR) est un PCR dont la progression est suivie à l'aide de lectures de fluorescence continue. La stratégie permettant la quantification est simple : sachant qu'à chaque cycle la concentration d'ADN_c double, on peut déduire la quantité d'ADN initialement présente dans un échantillon en mesurant le nombre de cycles de PCR nécessaire pour que la fluorescence liée à l'amplification d'ADN se démarque du bruit de fond. Ce nombre de cycles est appelé *cycle de quantification* (C_q). Le lien entre le C_q et la quantité d'ADN est établi à l'aide d'une courbe standard.

2.1.3. Variantes du PCR

Transcriptase inverse qPCR (RT-qPCR)

Pour effectuer un qPCR, il est aussi possible d'utiliser de l'ARN à condition que l'on transforme cet ARN en ADN à l'aide de la transcriptase inverse. On appelle cette technique *Reverse Transcriptase* qPCR. L'intérêt principal du RT-qPCR est que l'on peut évaluer l'expression des gènes des cellules ou des microorganismes. On peut aussi faire le RT-qPCR afin de quantifier seulement les microorganismes viables puisque l'ARN est instable et est produit par les cellules viables seulement [Dolan *et al.*, 2009; Fey *et al.*, 2004].

Multiplex PCR

Il est possible de rechercher plus d'une cible génomique dans un qPCR en utilisant différentes combinaisons d'amorces et de sondes [Doral, 2006]. Cela peut servir par exemple à détecter plusieurs souches de microorganismes dans un seul test [Jimmenez et al. 2011; Farajnia et al., 2008]. Cela peut s'avérer utile quand on recherche par exemple un nombre restreint de microorganismes pathogènes indésirables (par exemple: *E. coli*, *S. aureus*, *Salmonella*, etc.)

2.2. Fluorescence et mélange réactionnel

2.2.1. Stratégie de fluorescence

Pour détecter l'ADN amplifié lors d'un PCR, il existe trois stratégies de fluorescence possible. On peut soit utiliser un colorant libre, un colorant lié à l'amorce ou encore une sonde [Dorak, 2006].

Le colorant libre

Le colorant libre, par exemple le SYBER Green® est un fluorophore qui s'attache à l'ADN double brin. Lorsque lié à l'ADN, le signal de fluorescence de la molécule colorante augmente de plusieurs ordres de grandeur. Le signal de fluorescence augmente ainsi exponentiellement à chaque cycle de PCR.

Le colorant-amorce

Le colorant amorce est une amorce liée à un fluorophore dont la fluorescence n'est possible que lorsque l'amorce s'est hybridée avec le brin d'ADN (ou l'inverse). Le système est basé sur l'éloignement ou le rapprochement d'un *reporteur* (fluorophore) et d'un désactivateur; le fluorophore émet de la lumière lorsqu'il est éloigné du désactivateur ou l'inverse selon le système. Ce phénomène est appelé transfert d'énergie par résonance de type Förster (FRET). L'avantage de cette stratégie est que l'on peut détecter plus d'un gène d'intérêt dans un échantillon en utilisant plusieurs colorants-amorce dont l'émission est à des longueurs d'onde différentes.

La sonde

La sonde est un oligonucléotide distinct de l'amorce qui s'hybride sur le brin d'ADN. Dans le cas de la sonde TaqMan, celle-ci s'hybride à l'ADN puis le *reporteur* est détaché de la sonde lorsque la polymérase passe sur celle-ci permettant ainsi au *reporteur* d'émettre de la lumière. Un autre exemple est la sonde *molecular beacon* qui éloigne le *reporteur* du désactivateur, lorsque lié à l'ADN. Il existe d'autres types de sondes comme les sondes de type *minor groove-binding* (MGB), de type *Locked nucleic acid* (LNA) et les sondes d'hybridation .

2.2.2. Éléments du mélange réactionnel

Le mélange réactionnel doit être porté à 95 °C pour séparer l'ADN double brin en deux fragments d'ADN simple brin. Le PCR ne serait pas possible sans l'utilisation d'une polymérase résistante à la chaleur, telle que la Taq polymérase par exemple. Le magnésium qui est un cofacteur de la polymérase doit aussi être ajouté pour que la synthèse ait lieu. De plus, des dinucléotides, matière première de la polymérisation de l'ADN, sont nécessaires [Doral, 2006].

2.3. Analyse de données

L'analyse de données dépend du type de quantification désirée. Il existe deux types de quantification : la quantification relative et la quantification absolue. La quantification relative réfère à la comparaison du gène d'intérêt dont l'expression est variable avec un gène de référence dont l'expression est invariable [Pfaffl, 2001]. La quantification relative est utilisée lorsque l'on évalue le niveau d'expression d'un gène par rapport à une population de cellule. La quantification absolue réfère généralement à l'utilisation d'une courbe standard pour mesurer la concentration d'un gène d'intérêt dans un volume d'échantillons précis. Il est cependant possible d'effectuer une quantification absolue sans utiliser de courbe standard, mais plutôt en se basant sur une calibration optique et sur une analyse numérique de la courbe de fluorescence [Rutledge, 2004]. Dans le cas de la détection de microorganisme dans des échantillons provenant d'eau ou de produits pharmaceutiques, la **quantification absolue** est utilisée. Le travail ci-présent ne s'intéresse qu'à la quantification absolue avec l'utilisation d'une courbe standard. Néanmoins, la quantification que l'on nomme absolu est relative sur certains points : courbe standard, condition expérimentale, réactifs, etc [Smith *et al*, 2009].

2.3.1. Détermination du cycle de quantification (C_q)

Dans le cas de la quantification absolue avec courbe standard, le but ultime du traitement de données en qPCR est de trouver le cycle de quantification (C_q). Cette valeur permet de déterminer le nombre d'unités d'ADN présentes dans l'échantillon en utilisant la courbe standard. Il existe différentes façons de déterminer cette valeur. La valeur du C_q pour une même quantité d'ADN varie largement d'une méthode de quantification à l'autre. Le principe du cycle de quantification est d'utiliser un point dans la courbe de fluorescence qui possède une abscisse constante entre les réactions de PCR provenant d'une même quantité d'ADN. Ce point varie selon la méthode utilisée.

Méthode du *Threshold*

Une façon simple de déterminer le C_q est de le fixer par rapport au moment où la fluorescence se démarque du bruit de fond. Un seuil de coupure, appelé *threshold*, est fixé en déterminant l'écart type de la ligne de base. Généralement, l'écart type moyen de plusieurs réactions pour les mêmes conditions sera utilisé. Le seuil de coupure est généralement égal à la fluorescence de la ligne de base de la réaction de PCR auquel on ajoute 10 fois l'écart type moyen calculé [Life Technologies Corporation, 2012]. Le C_q correspond alors au cycle où la fluorescence mesurée atteint celle du seuil de coupure.

Méthodes basées sur la différentiation

Les maximums des dérivés possèdent généralement des abscisses constantes pour une même quantité d'ADN, on peut donc déterminer un C_q par rapport à une dérivée. La première dérivée correspond au point d'inflexion alors que la seconde dérivée correspond à l'accélération maximale de la fluorescence. Généralement, la différentiation de première ou de deuxième ordre est utilisée. La différentiation des données brutes génère des C_q imprécis en raison du bruit et de l'absence d'information entre les cycles. Ainsi, les méthodes utilisant les dérivées modélisent d'abord la courbe d'amplification [Spiess et al., 2008] ou lissent [Trampuz et al., 2006] celle-ci avant de déterminer les C_q .

La Figure 2.2 représente les courbes typiques de fluorescence de qPCR pour une expérience réelle. Une inspection visuelle de cette première courbe de fluorescence permet d'établir une première estimation du maximum de la seconde dérivée. Le calcul du C_q par un algorithme développé dans le cadre de ces travaux donne 26,9; 30,7 et 33,9 ce qui précise l'estimation visuelle.

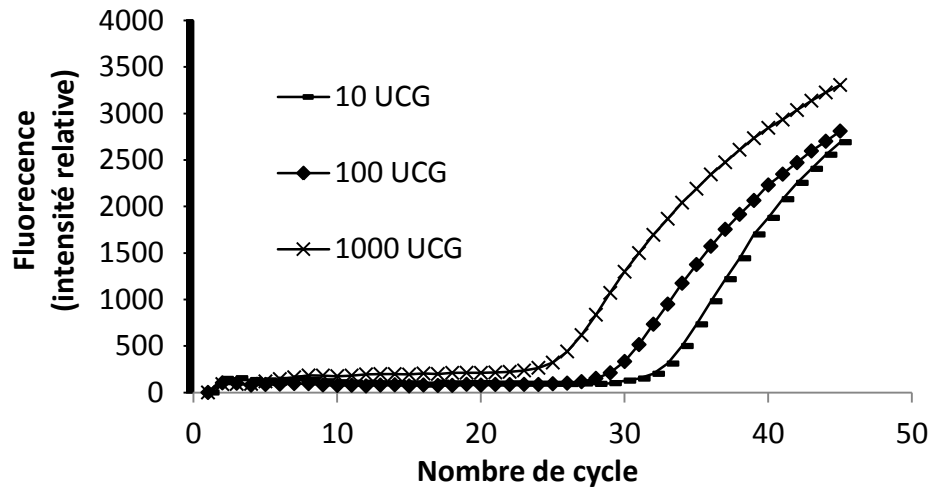


Figure 2.2. Courbes typiques d'un PCR

Autres méthodes

Outre le seuil de coupure et la différentiation, il existe plusieurs méthodes permettant de générer des C_q . Par exemple, la méthode du C_{y0} utilise l'abscisse de la tangente du point d'inflexion de la courbe modélisée [Guescini et al., 2008; Guescini et al., 2013]. La méthode FPLM utilise, quant à elle, le croisement de la phase exponentielle modélisée avec le seuil de coupure [Tichopad et al., 2003]. Cependant, d'autres méthodes mathématiques pour évaluer le cycle de quantification existent : CAMper-DART, LinRegPCR, PCR-Miner [Ruijter et al. 2013]. De plus, plusieurs méthodes déterminent la quantité d'ADN en utilisant la fluorescence initiale, F_0 , théorique calculée au lieu d'utiliser directement le C_q ; ces méthodes requièrent un calcul précis de l'efficacité des réactions de PCR pour chaque cycle.

2.3.2. Modélisation de la courbe sigmoïde

Pour plusieurs méthodes de quantification la courbe brute d'amplification d'ADN est modélisée. La modélisation de la courbe de PCR permet de déterminer les C_q ainsi que d'autres valeurs comme l'efficacité d'amplification par exemple. Le principal attrait de la modélisation est de réduire le bruit de mesure de fluorescence. Le modèle le plus connu est le modèle de la courbe sigmoïde à quatre paramètres. Cette méthode est basée sur la régression non linéaire des données de fluorescence. Les quatre paramètres sont illustrés dans la Figure 2.3. Le modèle de Tichopad [Tichopad et al., 2003], équation 2.1, est une expression du modèle à quatre paramètres. Le modèle à quatre paramètres peut être représenté différemment [Liu et Saint, 2002] et il est possible de lui ajouter un cinquième paramètre pour prendre en compte l'asymétrie [Spiess *et al.*, 2008].

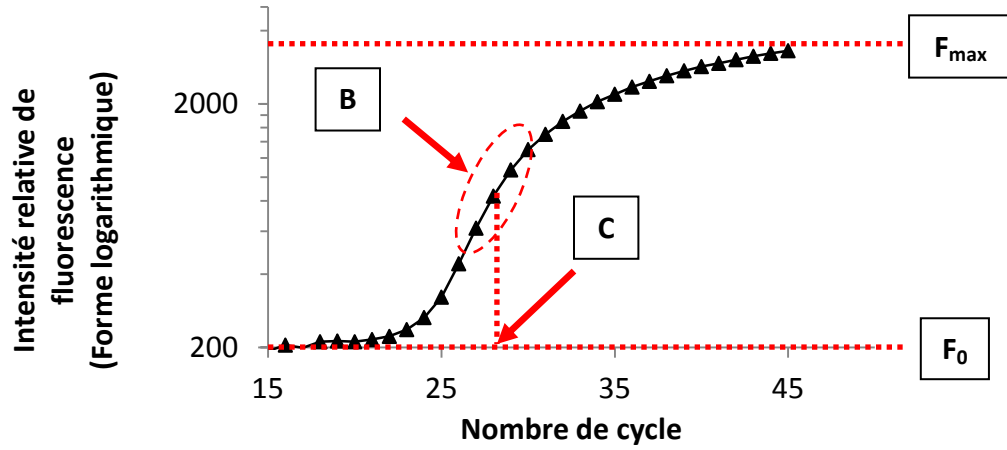


Figure 2.3. La courbe de fluorescence décrite par quatre paramètres

- F_0 est l'asymptote de fluorescence minimale, soit la ligne de base (*baseline*)
- B décrit la pente au point $F = (F_{max} - F_0) / 2$
- C est le cycle où $F = (F_{max} - F_0) / 2$
- F_{max} est l'asymptote de fluorescence maximale

$$F(n) = F_0 + \frac{F_{max} - F_0}{1 + \left(\frac{n}{c}\right)^{-b}} \quad 2.1$$

Outre la modélisation géométrique de la courbe sigmoïde, il existe d'autres types de modèle qui sont par exemple basés sur la mécanique d'amplification du PCR [Boggy et Woolp, 2010; Carr et Moore, 2012].

2.3.3. Détermination de l'efficacité d'amplification (E)

Le cas idéal où l'efficacité d'amplification est de 100 % signifie que la quantité d'ADN double à chaque cycle d'amplification. Expérimentalement, l'ADN est amplifié au maximum de son efficacité pour les premiers cycles d'amplification. Au-delà d'un seuil correspondant au maximum de la seconde dérivée, l'efficacité d'amplification diminue à chaque cycle. Le maximum de l'efficacité d'une réaction de PCR est dû à la chimie de l'échantillon (i.e. réactif de PCR, concentration d'ADN, présence d'inhibiteur ou d'activateur). La baisse d'efficacité de l'amplification durant la réaction est due à l'augmentation du rapport des produits sur les réactifs. Ainsi, afin de quantifier avec précision l'ADN d'un échantillon, l'efficacité d'amplification de PCR doit être incluse dans les calculs. Dans le cas de la quantification absolue avec courbe standard, l'efficacité est déterminée à l'aide de cette courbe standard. L'efficacité d'amplification ainsi obtenue est une moyenne des amplifications de l'ensemble des courbes utilisées pour réaliser la courbe standard, des premiers cycles jusqu'au C_q . Pour la quantification relative et la quantification absolue sans courbe standard, l'efficacité est déterminée en analysant la courbe d'amplification modélisée. Ces deux techniques de quantification ne font toutefois pas l'objet des présents travaux. Le lecteur peut respectivement consulter les travaux de Michael Pfaffl ou de Robert Rutledge pour plus d'information au sujet de ces techniques [Pfaffl, 2001; Rutledge et al, 2003; Rutledge, 2004].

2.3.4. Courbe standard et calcul de la quantité d'ADN

Le lien entre le cycle de quantification, C_q et la quantité d'ADN, en tenant compte de l'efficacité d'amplification et des particularités optiques du système, est établi à l'aide de la courbe standard. Une régression linéaire du logarithme des différentes concentrations d'ADN standard contre les C_q obtenus permet de générer la courbe standard. L'équation obtenue, l'équation 2.1, peut être réarrangée pour donner l'équation 2.3 qui est plus pratique.

$$C_q = m * \log_{10}[ADN] + b \quad 2.2$$

$$[ADN] = 10^{\frac{C_q - b}{m}} \quad 2.3$$

où b est l'ordonnée à l'origine et m la pente de la courbe standard. Cette courbe permet ainsi de quantifier l'ADN inconnue provenant de courbes d'amplification de PCR réalisées avec les mêmes réactifs et la même méthode.

2.3.5. Analyse statistique

Les données de qPCR, comme plusieurs types de données provenant d'échantillons biologiques sont sujettes à une variance importante. C'est pourquoi les données de qPCR doivent être interprétées à l'aide de statistiques de base. L'écart type (ET) ou l'écart type relatif (ETR), aussi appelé coefficient de variation (CV), ainsi que l'intervalle de confiance doivent accompagner les données. À cause de la nature logarithmique du lien existant entre la concentration d'ADN et le C_q il est déconseillé d'utiliser l'ETR avec le C_q [Bustin *et al.*, 2009] pour parler d'une variation de la quantité d'ADN (Tableau 2.1). Il est plutôt conseillé d'utiliser l'ETR calculé sur les variations de la quantité d'ADN.

Tableau 2.1. Relation C_q , [ADN] et ETR

	C_q [ADN]	
Tripliquats	26	2150
	25	4330
	27	1070
Moyenne	26	2520
ETR	3,8 %	65,9 %
m (pente)	-3,3	
b (ordonné)	37	

Lorsque les moyennes sont comparées, le test ANOVA suivi du test de Tukey HSD est utilisé pour évaluer si leur différence sont statistiquement significatives puisque l'écart-type est généralement élevé. Puisque la variance augmente proportionnellement à la moyenne et qu'elle n'est pas toujours uniforme dans les données, une transformation mathématique pour la stabiliser (ex. : transformation Box-Cox [Osborne, 2010]) pourrait être utile avant d'effectuer le test ANOVA. La façon adéquate de présenter des comparaisons de tests de qPCR est de les porter en graphiques de points, de type Box & Whisker, ou encore de représenter la moyenne et l'écart-type [Dorak, 2006].

2.4. Contrôle de la qualité

2.4.1. Analyse des courbes de PCR

La courbe de PCR typique se divise en quatre parties distinctes, telles que représentées dans la figure 2.4.

1. Ligne de base : amplification d'ADN indétectable par l'appareil
2. Phase exponentielle : amplification détectable où l'efficacité du PCR est maximale
3. Phase linéaire : amplification avec efficacité décroissance
4. Plateau: amplification nulle ou minimale due à l'épuisement des réactifs

Au niveau de la ligne de base, on s'attend à ce qu'elle soit constante. Une tendance à la baisse ou la hausse biaisera les résultats en les surestimant ou en les sous-estimant. Au niveau de la phase linéaire, on s'attend à ce que les portions linéaire de chaque courbe soient parallèles [Hugget et Bustin, 2011] entre elles, tel que représenté dans la figure.

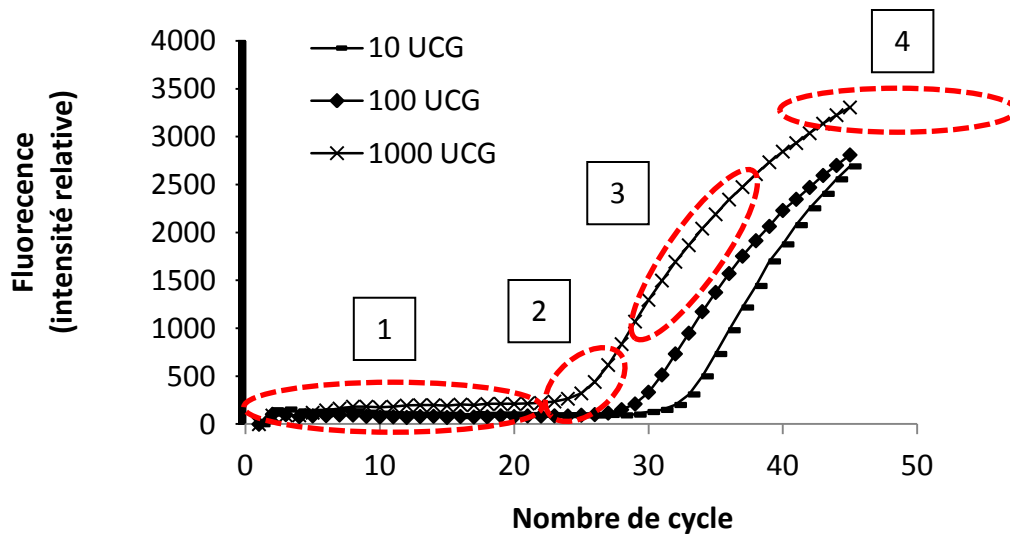


Figure 2.4. Division de la courbe de PCR

2.4.2. Analyse de la courbe standard

Pour une méthode de qPCR qui a été caractérisée, plusieurs courbes standard ont été réalisées ce qui permet de connaître la plage des valeurs acceptables pour la pente et l'ordonnée à l'origine de cette courbe. Ainsi, les courbes standard individuelles qui sont réalisées dans le cadre des essais sont évaluées par rapport à ces deux paramètres. Une déviation hors de ces plages permet de détecter des problèmes au niveau de la méthode (i.e. changement chimique des réactifs). Une courbe standard bien construite permet d'obtenir des résultats, en unité d'ADN, d'une précision allant 5 à 20 % d'écart type relatif [Rutledge et Côté, 2003] entre les répliques.

Efficacité réactionnelle : Analyse de la pente

La qualité du mélange réactionnel de PCR peut être jugée par l'efficacité de la mécanique d'amplification d'ADN (3), représenté par l'équation 2.4.

$$[\text{ADN}] = [\text{ADN}]_0 * (1 + E)^c \quad 2.4$$

où c est le nombre de cycles de PCR, E est l'efficacité d'amplification et $[\text{ADN}]_0$ est la concentration d'ADN présente initialement. L'efficacité absolue représente le facteur par lequel l'ADN est augmenté de façon exponentielle. Ainsi, pour une efficacité de 100 %, l'ADN double à chaque cycle de PCR, alors $E = 1$. L'efficacité en pourcentage est calculée à partir de la pente de la courbe standard (21) selon l'équation suivante :

$$E = \left(10^{\frac{-1}{m}} - 1 \right) \times 100\% \quad 2.5$$

où m est la pente de la courbe standard. Ainsi, un mélange réactionnel idéal donnerait une pente de -3.32 pour la courbe standard (voir la démonstration à l'annexe A. L'efficacité (E) généralement acceptée est de **85 à 110 %** [Postollec *et al.*, 2011].

Limite de détection du système : Analyse de l'ordonnée

L'ordonnée représente le nombre de cycles nécessaires pour détecter une seule molécule d'ADN. L'ordonnée dépend de la limite de détection de l'instrument par rapport au marqueur fluorescent [Dorak, 2006], de la méthode de quantification utilisée et de l'efficacité. Dans un système utilisant une méthode bien caractérisée, l'ordonnée attendue est connue et une plage d'ordonnée acceptable a été établie.

2.4.3. Analyse des contrôles

L'utilisation de contrôles permet de valider ou d'invalider le test. Ils sont aussi utilisés pour la résolution de problèmes, lorsqu'applicables. Le **contrôle positif**, permet d'assurer qu'il n'y pas d'inhibiteurs de PCR dans la préparation d'ADN. On considère qu'il y a inhibition lorsque le C_q de ce contrôle est retardé d'un cycle ou plus [Postollec *et al.*, 2011]. Ce contrôle devrait être en concentration proche de la limite de détection afin d'assurer la validité du test à la limite de celui-ci [Espy *et al.*, 2006]. Concrètement, ce contrôle est une séquence d'ADN différente de celle recherchée et sera amplifié indépendamment de la cible. Le **contrôle négatif**, aussi appelé contrôle de faux négatif, permet d'assurer qu'il n'y a pas d'amplification non spécifique; celui-ci ne contient pas de polymérase. Un contrôle de la préparation d'ADN, aussi appelé **contrôle environnemental**, permet d'évaluer le bruit de fond dû à la préparation d'ADN ou les contaminations liées à cette préparation; concrètement, c'est un contrôle qui a subi les étapes d'extraction d'ADN, mais qui ne contient pas l'échantillon initial; l'eau stérile de grade biologie moléculaire est un exemple de contrôle environnemental. Finalement, un contrôle de PCR, appelé *non-template control*, que nous traduirons en **contrôle sans ADN** permet d'évaluer le bruit ou la contamination produite lors de l'étape d'amplification.

2.4.4. Analyse des répliques

Selon la littérature, la différence entre les cycle de quantification (C_q) de répliques instrumentaux devrait être inférieur à 0,5 [Dorak, 2006].

3. DÉCOMPTE PAR QPCR — ÉTAT DE L'ART

3.1. Utilisation d'une sonde universelle pour les bactéries

Afin de faire le décompte de toutes les bactéries présentes dans un échantillon, la sonde et les amorces utilisées doivent être capables de s'hybrider sur le génome de n'importe quelle bactérie, on parle donc d'une sonde universelle. Puisque l'ADN ribosomal 16s est très préservé entre les organismes, c'est souvent une cible de choix pour la construction d'un ensemble d'amorces et de sonde [Nadkarni *et al.*, 2002]. Il existe aussi d'autres séquences du génome pouvant être ciblé comme *gyrB* [Yamamoto et Harayama, 1995], *tuf* [Paradis *et al.*, 2005; Tanaka *et al.*, 2010] ou *rpoB* [Dahllof, 2000].

La construction d'une courbe standard avec l'ADN d'un organisme en particulier (ex. : *E.coli*) pour la quantification de toutes les bactéries peut avoir comme effet de sous-estimer ou de surestimer certaines espèces. Ceci est dû au fait que la cible des amorces et de la sonde peut être présente en nombre variable d'une espèce à l'autre dans le génome de ceux-ci. Dans les expériences menées par Mangala A. Nadkarni, on constate qu'une courbe standard faite avec de l'ADN de *E.coli* ayant comme cible l'ADN ribosomal 16S sous estime *S.aureus* d'environ 50 % alors qu'une espèce comme *Prevotella melaninogenica* est surestimé de 1200 % [Nadkarni *et al.*, 2002]. Dans les expériences menées par Yuichiro Tanaka, un système de sonde universelle basée sur le gène *tuf* donne une meilleure quantification de l'ADN microbien. Le gène *tuf* étant présent à la hauteur de 1 ou 2 copies par génome bactérien entre les différentes espèces [Tanaka *et al.*, 2010]. Le gène *rpoB* serait également une bonne cible puisqu'il est présent en une seule copie dans le génome des différentes espèces [Dahllof *et al.*, 2000].

3.2. Équivalence entre le qPCR et le décompte sur pétri

Un décompte bactérien sur pétri est très différent d'une quantification bactérienne par qPCR. Les deux méthodes ont leurs avantages et leurs inconvénients. Il est donc nécessaire d'évaluer la réponse de ces deux méthodes par rapport à un même échantillon si on désire remplacer la méthode traditionnel par la nouvelle méthode. Les différences entre les deux méthodes proviennent de la variable mesurée qui n'est pas la même. Dans le cas du décompte sur pétri, on compte les bactéries qui sont capables de pousser sur un milieu de culture solide. De plus, ce sont les colonies qui sont comptées. On suppose qu'une colonie sur boîte de pétrie origine d'une seule cellule. Cependant lorsque des bactéries agrégées se retrouvent sur boîte de pétri, une seule colonie sera comptée [Sutton, 2012]. En somme, le décompte sur milieux solide ne permet pas de compter les bactéries viables non cultivables et est biaisé par les bactéries agrégées.

Dans le cas du qPCR, la variable mesurée est un fragment d'ADN présent dans le génome bactérien. Le qPCR mesure l'ADN des bactéries vivantes ou mortes et l'ADN génomique libre. Si une étape de filtration des cellules est incluse dans le protocole, l'ADN libre ne sera pas quantifié. Le nombre de fois où la cible génomique apparaît dans le génome, ainsi que l'ADN utilisé pour faire la courbe standard peut aussi influencer la quantification. En bref, la quantification par qPCR est biaisée par les bactéries mortes, la cible génomique et l'ADN utilisé pour faire la courbe standard.

Généralement, la quantification par qPCR donne des valeurs supérieures au décompte sur boîte de pétri, Tableau 3.1 . La fonction mathématique exprimant la mesure par qPCR en fonction de la mesure sur milieu solide varie largement d'un domaine à l'autre et le niveau de corrélation aussi, passant de $R^2 = 0,46$ [Haugland *et al.*, 2005] à $R^2 = 0,98$ [Hierro *et al.*, 2007].

Tableau 3.1. Équivalence entre le qPCR et la méthode conventionnelle dans la littérature

Domaine	Microorganisme ciblé	Plage étudiée (UFC)	Régression qPCR vs méthode traditionnelle (X = log UFC, Y = log UCG)	Source
Alimentaire, vin	<i>H. uvarum</i>	10 ³ à 10 ⁶	Y = 0.90X + 0.74; R ² = 0.98; N=12	*
	<i>S. cerevisiae</i>		Y = 1.11X - 0.81; R ² = 0.97; N=12	
Alimentaire, viande	Bactéries totales	10 ² à 10 ⁷	Y = 1,11Y - 0,60; R ² = 0.88, N=18	**
Clinique	<i>M. tuberculosis</i>	10 ² à 10 ⁴	R ² = 0,67	***
	<i>M. a. avium</i>		R ² = 0,91	
	<i>M. a. paratuberculosis</i>		R ² = 0,83	
Environnementale	<i>E. coli</i> Edgewater Beach 2006	0 à 10 ⁴	Y = 1,30X + 0,60; R ² = 0.88; N = 35	****
	<i>E. coli</i> Edgewater Beach 2007		Y = 0,79X + 0,49; R ² = 0,54; N= 27	
	<i>E. coli</i> Villa Angela Beach 2006		Y = 1,05X + 1,50; R ² = 0,59; N = 36	
	<i>E. coli</i> Villa Angela Beach 2007		Y = 0,70X + 1,50; R ² = 0,69; N = 38	
Environnementale	<i>Enterococcus ssp</i>	0 à 10 ⁴	Y = 0,53X + 1,56; R ² = 0,46	*****

* Hierro *et al.*, 2007** Dolan *et al.*, 2009*** Pathak *et al.*, 2007

**** Kephart et Bushon, 2009

***** Haugland *et al.*, 2005

Il est possible d'éliminer le biais apporté par les bactéries mortes sur la quantification du PCR en utilisant l'ARN au lieu de l'ADN [Dolan *et al.*, 2009]. Puisque l'ARN résulte de l'expression des gènes et qu'il est instable, il ne peut que signaler la présence de bactérie viable. Cependant, la préparation d'échantillon comporte plus d'étapes puisque l'ARN doit être transformé en ADN_c dans le processus de RT-qPCR.

3.3. Limite de détection et limite de quantification

La limite de détection ainsi que la limite de quantification du qPCR sont très variables selon la littérature consultée. Postollec (2011) publie dans sa revue de littérature, un tableau des limites de détection et de quantification dans différentes matrices alimentaires. On constate que la LOD varie entre 10^0 et 10^4 microorganismes par ml et que la LOQ varie entre 10^1 et 10^5 microorganismes par ml. De manière générale, la LOQ et la LOD se situent entre 10^2 et 10^3 organismes par ml [Postollec *et al.*, 2011]. La procédure d'extraction d'ADN, la matrice de l'échantillon, ainsi que la présence d'ADN contaminant explique cette variation. Une méthode de qPCR utilisant une sonde universelle pour la quantification de l'ADN total bactérien sera plus affectée par la présence d'ADN contaminant qu'une méthode de qPCR traditionnelle axée sur la quantification d'une espèce précise [Denoya, 2009]. L'utilisation d'une filtration permet aussi de réduire la LOQ et LOD en éliminant l'ADN extracellulaire [Urso *et al.*, 2009].

En ajoutant une étape de croissance bactérienne, ce qui rend le test qualitatif et non quantitatif, on peut obtenir une sensibilité de 1 UFC [Jimenez, 2011; Farajnia *et al.*, 2008]. Cette étape de croissance consiste à incuber l'échantillon dans un milieu de croissance contenant des nutriments et des éléments permettant d'inhiber l'effet bactériostatique de certains produits contenu dans les poudres pharmaceutiques. Cette stratégie permet pour les tests d'absence/présence d'utiliser un temps d'incubation beaucoup plus court que ce qui est utilisé pour les tests traditionnels; on parle d'un temps de cycle de 24-48h comparativement à 3 à 5 jours. Cette stratégie peut aussi être jumelée à la méthode du nombre le plus probable [United States Pharmacopeia, 2012] ou *most probable number method* (MPN) dans le but d'obtenir une méthode quantitative. Cependant, la méthode MPN est plus coûteuse en terme de réactif et de temps de manipulation.

4. TECHNOLOGIE ET MÉTHODE À L'ESSAI

4.1. Description de la méthode

4.1.1. Vue d'ensemble de la méthode

De l'échantillonnage aux résultats, quatre étapes majeures sont requises pour quantifier l'ADN : l'extraction, la concentration, l'amplification et la quantification. Premièrement, un échantillon liquide est passé à travers un filtre de 0.45µm pour récupérer les bactéries. L'ADN extracellulaire est théoriquement éliminé à cette étape alors que l'ADN intracellulaire est conservé puisqu'elle est emprisonnée à l'intérieur de la cellule retenue par le filtre. Ensuite, la membrane cellulaire est lysée et l'ADN est récupéré. Puis, l'ADN est concentré et purifié. Ensuite, l'ADN purifié est mélangé avec un ensemble de produits chimiques requis pour réaliser une réaction de PCR, communément appelé master mix. Ce mélange est placé dans un thermocycleur où la réaction de PCR aura lieu et où des lectures de fluorescence seront réalisées à chaque cycle d'amplification. Finalement, les données de fluorescences générées seront analysées par un algorithme dont la variable de sortie est le cycle de quantification (C_q). Ce C_q est ensuite converti en un nombre d'unités de copie de gène (UCG) en utilisant une courbe standard. Cette unité est différente de l'unité formant une colonie (UFC), mais y est proportionnelle.

Cette méthode utilise une sonde universelle, telle que décrite dans la section 3.1, ce qui permet la quantification de toute bactérie. Elle vise à remplacer le test du compte total de bactérie (TAC). Les travaux effectués dans les laboratoires de Pfizer Montréal ont été réalisés à l'aide d'un thermocycleur prototype ainsi qu'une procédure manuelle d'extraction et de purification d'ADN. Le prototype ainsi que la méthode sont en cours de développement par un fournisseur dont l'identité demeurera confidentielle pour des raisons contractuelles.

4.1.2. Particularité de la méthode

Appareil de PCR utilisé

L'appareil utilisé est une plateforme qui combine un thermocycleur, un système d'acquisition de données de fluorescence, un ordinateur, un moniteur et un lecteur de code-bar. L'ordinateur comporte un logiciel d'exploitation qui permet d'amorcer l'étape d'amplification, de faire la quantification, de voir les résultats et de les exporter. Le logiciel contrôle les cycles d'amplification avec des paramètres déjà établis par le fournisseur. Ce même logiciel comporte un algorithme qui analyse chaque courbe de fluorescence pour effectuer une quantification. Le lecteur de code-bar est utilisé pour associer le lot de réactif de PCR, ainsi que le lot de plateau, à la bonne courbe standard.

Plateau et puits de PCR

L'ADN purifié est placé dans le puits d'un plateau à PCR. Chaque échantillon doit être placé dans trois puits sur le plateau, conformément à la méthode. Deux puits sont utilisés pour l'amplification du gène d'intérêt, formant ainsi des duplicatas. Le troisième puits, ne contient pas de polymérase; c'est un contrôle négatif d'amplification d'ADN. Les puits utilisés pour les duplicatas comportent aussi une séquence d'ADN prête à être amplifiée avec un système marqueur fluorescent différent. Cette séquence d'ADN sert de contrôle positif puisque le patron d'amplification est toujours le même en absence d'inhibiteurs ou activateurs de PCR. Les plateaux de PCR sont configurés de façon à empêcher la fuite d'ADN du plateau vers l'extérieur, prévenant ainsi la contamination du laboratoire par l'ADN qui a été amplifié.

Contrôles de contaminations

En plus des contrôles d'amplification (positifs et négatifs) présents pour chaque échantillon, chaque expérience comporte deux contrôles de contaminations. Le premier consiste à traiter de l'eau stérile comme un échantillon de produit à analyser. Ce contrôle permet de détecter les contaminations pouvant provenir des manipulations, des réactifs, de la verrerie ou de l'environnement (comme d'une hotte contaminée par la manipulation de particules volatiles par exemple). Le second consiste à utiliser le tampon pour ADN comme un échantillon d'ADN prêt à être amplifié. Ce contrôle, appelé *non-template control* ou contrôle sans ADN, permet de détecter les contaminations provenant de la préparation des puits et du matériel utilisés pour

l'amplification d'ADN. Les contaminations peuvent provenir des réactifs, du plateau à puits PCR, du thermocycleur, etc.

4.1.3. Unité de quantification : UGC versus UFC

La méthode traditionnelle utilise l'unité formant une colonie (UFC) comme unité représentant l'abondance de cellule bactérienne dans l'échantillon. Une UFC représente l'unique cellule ou l'agrégat de cellule qui a poussé sur de l'agar-agar et qui est visuellement différenciable des autres colonies. À l'opposé, l'unité de copie de gène (UCG), unité de la méthode à l'essai, est représentative du nombre de copies d'un gène recherché dans un échantillon. La méthode permet de normaliser l'UCG pour que 1 UCG soit en moyenne équivalent à un génome bactérien sachant que le nombre de copies de ce gène dans le génome varie légèrement d'une espèce à l'autre. Ainsi, pour une seule cellule, le nombre d'UCG sera variable en fonction de l'espèce, mais sera en moyenne égale à un. Donc, puisque le nombre de cellules à l'origine d'un UFC varie à cause des agrégats et que le nombre d'UCG par cellule varie selon les espèces, le ratio UFC/UCG sera variable entre les espèces, mais devrait être constant pour une espèce si les conditions d'agrégation restent constantes. Néanmoins, il est expérimentalement difficile de conserver un ratio UCG-UFC constant si l'état métabolique des cellules est variable (cellule latente, versus cellule en division), voir section 6. Les deux types de comptes sont donc des approximations de la charge réelle en cellule vivante. Une méthode précise de compte de cellules vivantes serait d'effectuer un compte au microscope à l'aide d'un marqueur de viabilité, comme par exemple le marquage *life/dead*, ce qui n'est pas pratique dans un environnement industriel. On comprend donc que les deux types de quantification ont des défauts qui doivent être pris en compte lorsque des résultats rapportent un compte d'UFC et d'UCG. La section 3.2 discute de la corrélation entre les UCG et les UFC de la littérature et illustre que la corrélation n'est pas parfaite dû aux raisons citées précédemment.

4.2. Procédure

4.2.1. Préparation des cellules

Dans le but d'évaluer la méthode de qPCR, des bactéries sont préparées pour être ajoutées aux échantillons d'eau stériles. La préparation de cellule permet premièrement d'introduire une concentration ciblée de bactérie dans l'échantillon et deuxièmement de comparer la méthode de qPCR à la méthode traditionnelle du décompte sur boîtes de pétrie. L'étape de préparation des cellules n'est utile que pour l'évaluation de la méthode.

Les stocks de bactéries sont conservés sur des pentes* d'agar réfrigérées à 4 °C. La procédure standard est de prélever un échantillon du stock de bactéries et de l'étaler sur une pente vierge qui sera incubée à 35 °C pendant 16 à 24 h. La pente fraîche* est ensuite grattée, lavée avec 9ml de saline (9g NaCl/litre), puis transférée dans un tube de verre. La solution bactérienne est ensuite diluée jusqu'à un facteur de dilution de 10E-8. La dilution 10E-8 est à une concentration qui avoisine les 10E2 UFC/ml, dépendamment de l'espèce. Ensuite, 200 µl de la dilution 10E-8 ou 10E-7 sont étalés sur l'agar d'une boîte de pétri, juste avant l'exécution du protocole d'extraction d'ADN. Puis, les dilutions bactériennes pertinentes sont utilisées pour les tests de la méthode de qPCR. Finalement, une certaine quantité de suspension d'une des huit dilutions est transférée par pipetage du tube vers un contenant de plastique contenant 100 ml d'eau purifié. L'échantillon de 100 ml d'eau contaminé par une quantité connue de bactéries sera utilisé pour mettre à l'épreuve la méthode de qPCR.

Il est à noter que pour les tests effectués en 2012, les bactéries provenaient toujours d'une pente fraîche. En 2013, la procédure a été changée et on utilisa une pente âgée de 7 jours. Ce changement provient des conclusions de tests ayant évalués l'impact de l'âge de la pente et de la suspension, chapitre 6. Pour de l'information, au niveau des terminologies des pentes d'agar fraîche ou âgée, se référé à la section 6.2.1.

4.2.2. Procédure détaillée

Extraction

L'échantillon liquide qui a été recueilli dans un récipient de 200 ml est d'abord filtré à l'aide d'un filtre de 0,45 µm relié à un tuyau; l'eau est poussée à l'intérieur du tuyau en utilisant une pompe péristaltique. Le filtre a la particularité d'être à l'intérieur d'un support hermétique, que nous appellerons porte-filtre, possédant une entrée et une sortie pouvant être refermées par des capuchons. Les cellules sont ainsi immobilisées sur la membrane du filtre. Ensuite, une solution de tampon de lyse est ajoutée au porte-filtre. Le porte-filtre est fermé et placé dans un sonicateur. Les cellules sont lysées par l'effet combiné de la sonication et du tampon de lyse. Ensuite, le porte-filtre est placé sur un bloc chauffant pour inactiver les enzymes qui pourrait dégrader l'ADN.

Purification

Une fois les cellules lysées, l'ADN est libéré et peut être purifié. Le porte-filtre est ensuite placé à l'entrée d'un filtre NanoSep™ de 30 kDa de « molecular weight cutt-off » (MWCO). La sortie du porte-filtre s'insérant dans l'entrée du filtre Nanosep. Les filtres emboîtés sont centrifugés à 2000g, ce qui fait passer l'ADN du porte-filtre au filtre Nanosep. Le filtre Nanosep est ensuite centrifugé à 6000g: le filtre conserve l'ADN et laisse les petites molécules traverser. De l'éthanol est ensuite ajouté au filtre qui est centrifugé afin de laver l'ADN une première fois. Ensuite, un deuxième lavage est effectué avec une solution tamponnée. Enfin, l'ADN purifié est remis en suspension avec un diluant tamponné et est ensuite pipetté dans un tube de PCR pour l'amplification d'ADN à venir.

Amplification

Une fois l'ADN purifié, il est ajouté dans les puits de PCR avec le *master mix*. Puis le support contenant les puits est apporté à l'appareil de PCR. Ensuite, le logiciel de l'appareil invite l'utilisateur à numériser à la fois le code-bar du mélange de réactif PCR et de la plaque à puits afin d'associer les numéros de lots avec la courbe standard qui lui est associée. Finalement, l'amplification d'ADN est initiée avec des paramètres que l'utilisateur ne peut pas modifier.

Quantification et courbe standard

Lorsque l'étape d'amplification est terminée, un algorithme analyse les données brutes de fluorescences pour en extraire le cycle de quantification (C_q) pour chaque échantillon. Le lien entre le C_q et les UCG, pour les conditions expérimentales utilisées, est établi par la courbe standard. Cette courbe standard a été réalisée en utilisant des dilutions d'ADN bactériens standard de 10, 50, 100, 1000 et 10 000 UCG. La régression linéaire utilisée entre le logarithme en base 10 des concentrations d'ADN utilisé et le C_q , suite à l'amplification, permet d'obtenir les coefficients de la courbe standard. Les résultats de la courbe standard sont utilisés par le logiciel de la plateforme de PCR pour calculer les UGC de chaque échantillon. Pour plus de détails concernant les principes de quantification et les courbes standards, consulter la section 2.3 ainsi que le chapitre 7.

5. PERFORMANCES ANALYTIQUES

5.1. Mise en contexte

Afin de statuer sur la capacité de la méthode à quantifier les bactéries dans un échantillon pharmaceutique, la méthode sera évaluée selon les normes pharmaceutiques de la « United State Pharmacopeia » (United State Pharmacopeia, 2012) concernant la validation de méthodes microbiologiques alternatives. La méthode doit démontrer qu'elle respecte les critères USP pour être validée et soumise aux instances réglementaires pour son utilisation dans le domaine pharmaceutique. Dans le cas de la technologie sous évaluation, trois chapitres de la réglementation USP sont pertinents :

« USP <61> *Examination of nonsterile products : microbial enumeration tests* » comporte des directives strictes concernant la préparation des milieux, la préparation des microorganismes, la promotion de croissance des milieux, les contrôles, la neutralisation des agents antimicrobiens et la méthodologie en général concernant la méthode traditionnelle.

« USP <62> *Microbiological examination of nonsterile products : tests for specified microorganisms* » discute des tests qui doivent être réalisés pour détecter des organismes pathogènes spécifiques. Les organismes pathogènes spécifiés dans USP<62> ont été utilisés pour l'évaluation de la technologie décrite dans les présents travaux.

« USP <1223> *Validation of alternative microbiological methods* » comporte des indications générales sur la validation de méthode microbiologique alternative. Les indications de ce chapitre en ce qui a trait au caractère analytique de la méthode n'ont pas nécessairement à être respectés dans leur intégralité étant donné que l'objectif de ce chapitre est surtout d'orienter les spécialistes en validation plutôt que des les contraindre. La validation de méthode alternative doit démontrer que la nouvelle méthode est meilleure ou équivalente à l'ancienne méthode au niveau du contrôle de la flore microbienne par quantification de celle-ci. Les travaux présentés ici évalueront la méthode au niveau des spécifications analytiques de USP <1223>. Les résultats devront ensuite être analysés par rapport à ces spécifications avec certaines réserves étant donnée la relation variable existante entre les UFC et les UCG . Les

conclusions finales quant aux performances de la nouvelle méthode ne pourront être réalisées à la fin de ce chapitre. Les différences fondamentales entre les deux méthodes seront étudiées dans les prochains chapitres ce qui permettra de conclure en tenant compte de certaines particularités qui débordent du caractère analytique.

5.2. Méthodologie

5.2.1. Exactitude — taux de récupération

L'exactitude représente la capacité de la méthode alternative à générer des résultats qui sont fidèles à ceux générés par la méthode traditionnelle. Cependant, dans le cas des méthodes qui ne sont pas basées sur la croissance pour former des colonies ou pour générer de la turbidité, on accepte que la méthode génère un nombre de microorganismes en apparence supérieur à la méthode traditionnelle. USP <1223> précise que la méthode devrait donner une estimation supérieure à 70 % de l'estimation fournie par la méthode traditionnelle. Ainsi, dans le cas présent, le critère d'exactitude n'est pas un critère d'exactitude instrumentale, mais plutôt un seuil, un taux de récupération, à atteindre relativement à la méthode de compte traditionnel. Un rapport entre les UCG et les UFC, appelé taux de récupération, supérieur à 0.7 est attendu pour les résultats des tests.

L'exactitude a été évaluée en utilisant des données déjà existantes provenant de tests exécutés sur plusieurs souches pour l'évaluation du critère de spécificité. Les concentrations bactériennes utilisées se situent entre 50 et 5000 UFC.

5.2.2. Spécificité — universalité

L'objectif du critère de spécificité est de démontrer que la méthode répond adéquatement au besoin du test. Dans le cas du test de TAC, le besoin est le dénombrement de toutes bactéries. Ainsi, la méthode doit être en mesure de quantifier les bactéries de toutes les souches présentes dans un échantillon. Donc, on évalue l'universalité de la méthode pour le règne bactérien. L'évaluation de ce critère n'est pas exhaustive étant donné le nombre limité de souches testées.

La méthode de qPCR et la méthode traditionnelle ont parallèlement quantifié des suspensions de quatre souches de bactéries pathogènes spécifiées par USP <62> , c'est à dire *Escherichia coli*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Salmonella thyphimurium*, ainsi que *Bacillus subtilis* et trois souches environnementales isolées dans le laboratoire de l'usine de Montréal.

Salmonella thyphimurium, *Bacillus subtilis*, *Micrococcus luteus*, *Staphylococcus hominis* et *Staphylococcus warneri* ont été testés en duplicata uniquement pour l'évaluation de ce critère. Les résultats pour les trois autres souches proviennent d'essais visant l'évaluation des autres critères. Ainsi, les résultats d'essais pour *Escherichia coli* et *Staphylococcus aureus* proviennent d'essais préliminaires d'analyse de précision et les résultats d'essais de *Pseudomonas aeruginosa* proviennent d'essais pour l'analyse de linéarité.

5.2.3. Précision

La précision est la mesure du niveau de variation pour les résultats provenant d'un même échantillon. Selon USP <1223>, l'écart type relatif (ETR) pour les répliques doit être inférieur à 35 %. La précision a été évaluée en utilisant des répliques provenant des tests de linéarité et des tests de robustesse. Cinq suspensions de *S. aureus* en réplique quadruple, quatre suspensions d'*E. coli* en réplique quadruple et huit suspensions d'*E. coli* en duplicata pour un total de 17 suspensions ont été utilisées pour le calcul de la précision.

5.2.4. Linéarité

La linéarité d'un test microbiologique quantitatif est sa capacité à produire des résultats proportionnels à la concentration de microorganismes présente dans l'échantillon pour une certaine plage [United States Pharmacopeia, 2012]. La régression linéaire des UCG en fonction des UFC permet d'évaluer la linéarité de la méthode. Étant donné que la variance des UCG, ainsi que celle des UFC est fonction de la concentration, une transformation logarithmique en base 10 des UCG et UFC est nécessaire pour stabiliser la variance des résultats. Le coefficient de détermination (R^2) de la régression du logarithme des UCG et UFC permet de statuer sur linéarité de la méthode. USP <1223> suggère que la méthode à l'essai démontre un R^2 supérieur à 0,95.

La linéarité de la méthode a été évaluée pour trois souches en triplicata sur cinq différentes concentrations. Les souches sont *Staphylococcus aureus*, *Escherichia coli* et *Pseudomonas aeruginosa*.

5.2.5. Plage et limites de quantifications

Plage de quantification

La plage de quantification est l'intervalle des concentrations pour lequel les bactéries sont quantifiées avec exactitude, précision et linéarité [United States Pharmacopeia, 2012]. Ainsi la plage est déterminée lorsque les limites inférieures et supérieures de quantification ont été calculées.

Limites de quantifications

La limite de quantification est définie en tant que limite à laquelle les bactéries sont encore détectées et quantifiées correctement en tenant compte de la précision et de l'exactitude. Au-delà de ces limites, les bactéries sont encore détectées, mais pas convenablement quantifiées. Les données utilisées pour trouver les limites de quantification proviennent de l'ensemble des tests effectués.

5.2.6. Robustesse interne

La robustesse interne est la capacité de la méthode à supporter des variations acceptables de certains paramètres du protocole, sans que les résultats en soient affectés de manière à poser un problème dans l'interprétation de ceux-ci. Ce critère a été évalué, au cours de l'étude, par trois différentes manières. Dans un premier temps trois facteurs qui étaient suspectés d'avoir un effet sur les résultats ont été testés de façon uni-variée. Ensuite, un plan expérimental a été monté pour analyser plusieurs facteurs pouvant avoir un impact sur les résultats. Puis finalement, une analyse des données historiques, a permis d'évaluer l'impact de certains facteurs.

Analyse univariée

Durant les débuts de l'étude, la variance entre les réplicas était élevée et des doutes avaient été soulevés sur certaines étapes du protocole. L'étape de filtration de l'échantillon et l'étape de re-suspension de l'ADN étaient suspectées de contribuer aux sources d'erreurs. Au niveau de la filtration de l'échantillon, sur 100 ml d'échantillons, environ 10 ml n'étaient pas filtrés pour des contraintes pratiques et on supposait que ces 10 ml pouvaient être plus concentrés que le reste de l'échantillon dû à la sédimentation des cellules. On pensait aussi que des cellules pouvaient rester adsorbées à la tuyauterie utilisée pour pomper le liquide du contenant de l'échantillon vers le filtre. Pour l'étape de re-suspension d'ADN, il avait été supposé que la technique de pipetage avait un rôle important.

Les différents traitements utilisés pour vérifier l'impact des facteurs nommé précédemment sont inscrits dans le Tableau 5.1. L'effet de la sédimentation sera vérifié en comparant le traitement A vs le traitement B. L'adsorption de cellules à la surface de la tuyauterie sera évaluée en comparant le traitement A au traitement C. L'effet du pipetage sur la re-suspension sera évalué en comparant le traitement C au traitement D.

Tableau 5.1. Traitement de l'essai uni-varié pour évaluer la robustesse

Traitement	Ajout de 1 ml de suspension bactérienne	Quantité filtrée	Pipetage par va et vien
A	Dilution dans 120 ml H ₂ O	120 ml / 120 ml	5X
B	Dilution dans 120 ml H ₂ O	100 ml / 120 ml	5X
C	Directement dans le filtre	100 %	5X
D	Directement dans le filtre	100 %	2X

Analyse avec plan d'expérience factoriel

Un plan factoriel a été monté pour évaluer l'impact des différentes étapes du protocole qui ont été identifiées, lors d'un travail de remue-ménages, comme étant critiques. Les étapes identifiées ont été analysées pour leur sensibilité aux variations expérimentales. Les étapes suivantes ont été soumises à des variations raisonnables: filtration, sonication, centrifugation et congélation d'ADN.

Au niveau de la filtration, la présence de bulle sur le filtre laissait supposer qu'une partie des cellules recueillies par le filtre ne seraient pas en contact avec le tampon de lyse, ce qui pourrait affecter l'efficacité de la lyse cellulaire et ainsi diminuer le taux de récupération. Pour recréer l'effet des bulles sur le filtre, 0,5 ml d'air a été introduit à l'intérieur de l'enceinte du filtre en utilisant une seringue. Au niveau de la sonication, on avait supposé qu'un délai entre l'ajout du tampon de lyse et la sonication pouvait avoir un impact sur les résultats; on a testé avec un délais de 60 minutes et sans délais. En effet, si les cellules se lysent prématurément, il est possible que l'ADN soit dégradé. Au niveau de la centrifugation, il était supposé qu'une centrifugation excessive pouvait engendrer la perte d'ADN; on a fait varier le temps de centrifugation entre 7 et 25 minutes. Au niveau de la congélation d'ADN, on cherchait à déterminer si la congélation engendrait la perte d'ADN. La congélation étant une étape optionnelle qu'on utilisait lorsque l'ADN purifiée ne pouvait subir l'étape d'amplification le jour même de l'extraction. Le plan expérimental est un plan factoriel fractionnaire à deux niveaux pour 16 essais réalisés en duplicata ($2^{4-1} \times 2$), voir Tableau 5.2. L'analyse ANOVA du plan permettra d'évaluer quels sont les facteurs qui influencent les résultats.

Tableau 5.2. Plan fractionnaire 2^{4-1} en duplicata pour évaluation de la robustesse

Ordre standard	Ordre aléatoire	Numéro de réplicat	Bulle	Délais sonication (minutes)	Centrifugation (minutes)	Congélation ADN
1	1	1	-	-	-	-
14	2	2	+	-	+	-
5	3	1	-	-	+	+
9	4	2	-	-	-	-
7	5	1	-	+	+	-
15	6	2	-	+	+	-
6	7	1	+	-	+	-
11	8	2	-	+	-	+
3	9	1	-	+	-	+
13	10	2	-	-	+	+
16	11	2	+	+	+	+
2	12	1	+	-	-	+
10	13	2	+	-	-	+
4	14	1	+	+	-	-
12	15	2	+	+	-	-
8	16	1	+	+	+	+

Analyse de données historiques : nettoyage de pièce réutilisable

Lors de l'étude, il a été supposé que les résultats pouvaient être sensibles aux contaminations croisées entre les expériences de PCR. En effet, des doutes avaient été émis par rapport à l'efficacité du nettoyage d'une pièce de l'équipement étant utilisé entre chaque expérience de PCR. Cette pièce avait le potentiel d'entrer en contact avec les échantillons d'ADN. Ainsi, on soupçonnait que les résultats pouvaient être influencés par la charge en ADN de l'expérience précédente. Pour éliminer ce problème potentiel, la procédure de nettoyage de cette pièce fut modifiée. La procédure originale demandait à ce que cette pièce d'équipement soit nettoyée avec de l'éthanol entre chaque séance de PCR, et suggérait de l'autoclaver chaque semaine. Le protocole a été modifié afin d'autoclaver cette pièce d'équipement entre chaque expérience de PCR.

Dans le but d'évaluer si l'étape de nettoyage de cette pièce a une influence sur la robustesse de la méthode, les résultats provenant de la période avant la modification du protocole et celle après la modification du protocole seront comparés. Plus spécifiquement, le niveau d'ADN retrouvé dans les contrôles de PCR sera comparé pour les deux périodes.

5.2.7. Sommaire des essais

Les essais réalisés pour évaluer les critères USP sont énumérés dans le Tableau 5.3. On constate que les tests ont été effectués sous différentes versions du protocole. Une première modification au protocole a été effectuée afin d'éliminer des problèmes d'inhibition de PCR. Puis, suite à l'analyse de l'impact de la préparation des cellules décrite au chapitre 6, à l'amélioration du nettoyage d'une pièce d'équipement particulière et à l'analyse de robustesse décrite dans la section 5.2.6, le protocole fut légèrement modifié. Ces modifications eurent un impact positif sur la précision. Ainsi, la précision fut mesurée avec les données du dernier protocole en vigueur et la linéarité fut testée à nouveau pour la souche dont le score était le plus bas.

Tableau 5.3. Sommaire des essais de l'évaluation analytique

Critères évalués	Version du protocole	Souches testées	Niveau de concentration testé	Échantillons par niveau	Échantillons totaux
Linéarité, LOQ, plage	Version 1	3 (<i>E.coli</i> , <i>P.aeruginosa</i> , <i>S.aureus</i>)	5	3	45
Précision (préliminaire)	Version 2	2 (<i>E.coli</i> , <i>S. aureus</i>)	3	5	30
Spécificité	Version 2	<i>B. subtilis</i> , <i>S. thyphimurium</i> , <i>M.luteus</i> , <i>S.hominis</i> , <i>S.warneri</i>	1	2	10
Robustesse #1	Version 2	<i>E. coli</i>	1	12	12
Robustesse #2	Version 3	<i>E. coli</i>	8	2	16
Linéarité, précision	Version 3	<i>S.aureus</i>	5	4	20

5.3. Résultats

5.3.1. Exactitude — taux de récupération

La méthode alternative doit quantifier au moins 70 % de l'énumération donnée par la méthode. Pour les différentes souches testées, le taux de récupération UCG-UFC varie de 0,7 à 5,6, voir le Tableau 5.4. L'estimation par qPCR est supérieure à 70 % de l'estimation de la méthode traditionnelle pour toutes les souches testées, satisfaisant ainsi le critère du taux de récupération. Le nombre d'UCG par cellule bactérienne varie d'une souche à l'autre, comme discuté dans la section 3.1. En plus des différences de nature génétique, des différences en lien avec le taux d'agrégation ont probablement une influence sur les résultats. La variation UCG-UFC provient du fait que les unités UCG et UFC sont deux unités différentes et sont une expression grossière de la quantité de microorganisme dans un échantillon, tel que décrit en détails à la section 4.1.3.

Tableau 5.4. Taux de récupération par la méthode qPCR

Souches bactériennes	n =	Ratio UCG/UFC
<i>Escherichia coli</i>	5	2,0
<i>Staphylococcus aureus</i>	5	0,8
<i>Pseudomonas aeruginosa</i>	12	0,7
<i>Salmonella thyphimurium</i>	2	2,5
<i>Bacillus subtilis</i>	2	0,7
<i>Micrococcus luteus</i>	2	1,4
<i>Staphylococcus hominis</i>	2	5,6
<i>Staphylococcus warneri</i>	2	1,1
Moyenne des souches	/	1,9

5.3.2. Spécificité — universalité

La technologie doit permettre la détection de toute souche de bactéries pour être en mesure de remplacer les tests TAC. Ainsi, la technologie a été mise à l'épreuve avec huit souches bactériennes représentant à la fois les bactéries pathogènes spécifiées indésirables par USP et les bactéries souvent identifiées dans l'environnement de l'usine. Les huit souches testées ont été détectées, donc les exigences de spécificité ont été atteintes pour l'analyse restreinte, voire le Tableau 5.4 pour la liste des microorganismes testés. Le nombre de souches testé est toutefois restreint pour des raisons pratiques. Le fournisseur a déjà démontré que son système fonctionne avec un large éventail de souches (résultats non-publiés).

5.3.3. Précision

Il est suggéré d'évaluer la précision à des concentrations proches des limites de quantification puisque la précision peut diminuer aux limites. Les essais précédents, réalisés avec un protocole non optimisé, ont montré qu'il n'était pas nécessaire de le faire pour cette méthode puisque la précision n'a pas été corrélée avec la concentration, voir Figure 5.1. Par conséquent, la précision a été évaluée à l'aide des données existantes.

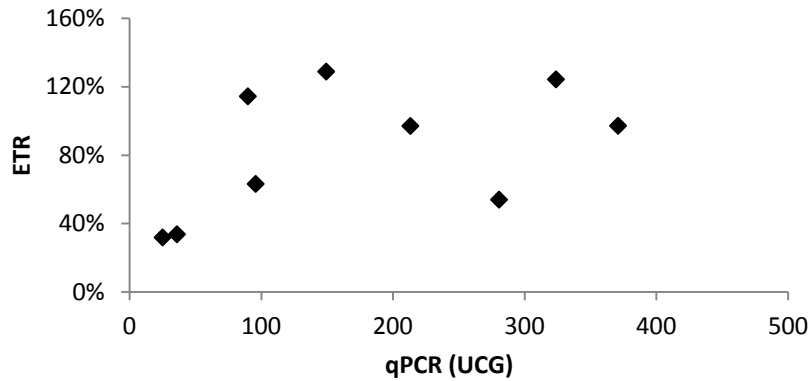


Figure 5.1. ETR pour la plage inférieure de concentration pour les essais de 2012

Les écarts types relatifs des 17 suspensions, analysées avec le protocole final de qPCR, sont présentés dans la Figure 5.2. La précision moyenne, donnée par l'ETR, était de 33 % avec un intervalle de confiance de 6 %. La précision respecte la norme USP (35 %).

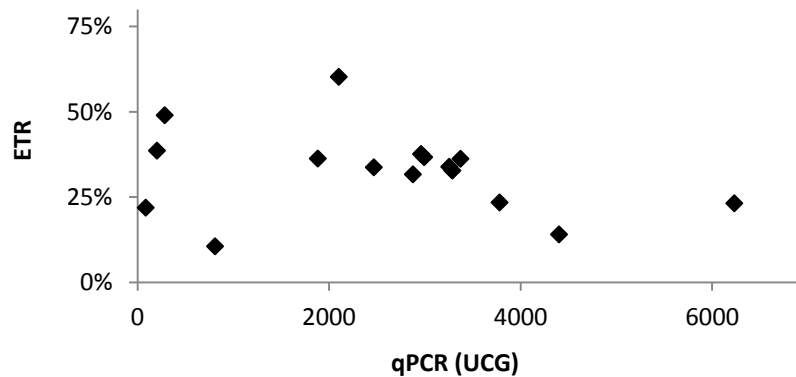


Figure 5.2. ETR pour les 17 suspensions réalisées avec le dernier protocole en vigueur

L'analyse des sources de variation a montré que 40 % de celle-ci provient de la réaction de PCR et de l'algorithme de quantification et que le 60 % restant provient du protocole. Étant donné que le protocole comporte de nombreuses étapes manuelles qui sont des sources potentielles d'erreurs, il est attendu que la variance provenant des étapes manuelles d'extraction et purification d'ADN devrait diminuer lorsque ces étapes seront automatisées par le fournisseur. Ainsi, la précision ne devrait pas être problématique lorsque la méthode développée par le fournisseur sera sur le marché.

5.3.4. Linéarité

Linéarité

La linéarité sera jugée en fonction de son R^2 , avec un objectif de 0,95. Les trois souches ont été testées en triplicatas pour des concentrations allant de 50 à 50 000 UFC, voir la Figure 5.3.

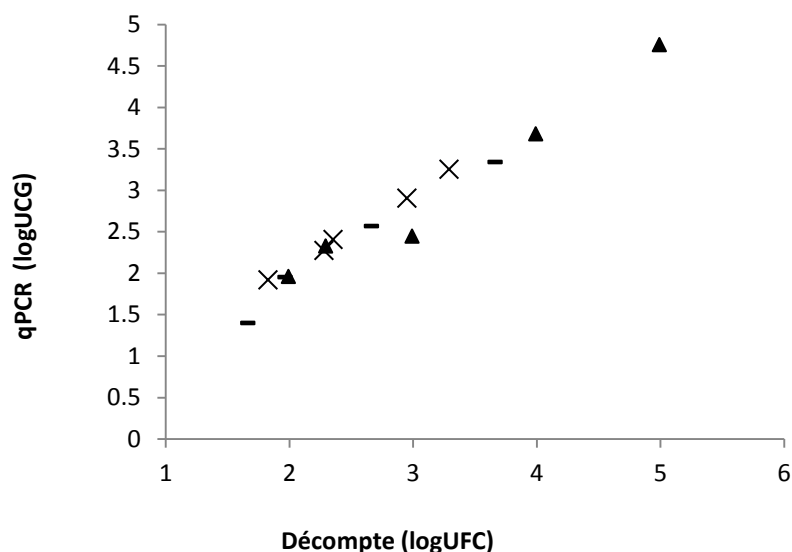


Figure 5.3. Tests de linéarité. *P. aeruginosa* (tiret), *E. coli* (triangle), *S. aureus* (croix)

Le coefficient de détermination (R^2) varie entre 0,90 et 1 et la pente de la régression a varié de 0,8 à 0,93, Tableau 5.5.

Tableau 5.5. Valeur de la régression linéaire

	R^2	Pente	Ordonnée
<i>S. aureus</i> 2013	1,00	0,91	0,24
<i>P. aeruginosa</i> 2012	0,98	0,93	0,01
<i>E. coli</i> 2012	0,90	0,81	0,52

Les valeurs de R^2 satisfont aussi aux exigences USP, sauf pour *E. coli*. La valeur de R^2 moyenne pour les trois souches est de 0,96 ce qui démontre l'atteinte de l'objectif. La valeur plus faible de R^2 obtenue pour *E. coli* est expliquée par le fait que le protocole utilisé n'était pas optimal lors des essais. La section 5.2.7. donne plus de détails sur l'évolution du protocole.

5.3.5. Plages et limites de quantification

Limite inférieure de quantification

La limite inférieure de quantification (LQ_I) est la plus petite valeur pouvant être quantifiée avec précision et exactitude [United States Pharmacopeia, 2012]. Comme montrée par la Figure 5.1, la précision ne diminue pas dans la plage inférieure des concentrations testées (50-500 UFC). Aussi, le taux de récupération, appelé exactitude par USP, ne diminue pas pour les faibles concentrations testées, exceptées pour *E.coli* qui a été testé avec un protocole non optimisé. Ainsi la LQ_I ne peut être basée sur la précision et l'exactitude. La LQ_I devrait plutôt être basée par rapport au bruit de fond puisque celui-ci ajoute un biais à la quantification. Plus la concentration testée est faible, plus le poids de ce biais devient important. Le signal du bruit de fond, donné par l'ADN des cellules provenant de l'environnement, des manipulations et des réactifs, est de 20 UCG avec un écart-type de 10 UCG. Ainsi, il serait raisonnable de proposer une limite inférieure de quantification équivalente au bruit de fond en y ajoutant trois fois son écart-type ce qui donnerait 50 UCG. Soustraire le bruit de fond de chaque résultat et instaurer une (LQ_I) de 50 UCG permettrait ainsi l'obtention de résultats fiables. La valeur de la LQ_I devrait être recalculée avec de nouvelles valeurs de bruit de fond si les conditions d'essais sont changées, telles que le lieu, les réactifs et le type d'échantillon par exemple.

La LQ_I et la limite supérieure de quantification (LQ_S) donnent la plage de quantification. La LQ_S est la limite supérieure pour laquelle la quantification est exacte et précise. Aucun test n'a été effectué spécifiquement pour déterminer cette valeur puisque parce que celle-ci se trouve au-delà des spécifications des produits de l'usine de Pfizer Montréal. La valeur de celle-ci n'a donc pas été trouvée avec précision. Cependant, les résultats permettent d'établir deux bornes entre lequel la LQ_S se trouvera assurément. Premièrement, pour la plage des tests effectués, qui est de 50 à 10 000 UFC, la précision et le taux de récupération sont indépendant de la concentration, Figure 5.2 et Figure 5.4. Donc, les bactéries peuvent être quantifiées avec exactitude et précision dans les plages de concentrations évaluées, ce qui permet de déterminer que la limite supérieure de quantification se trouve au-delà de cette plage.

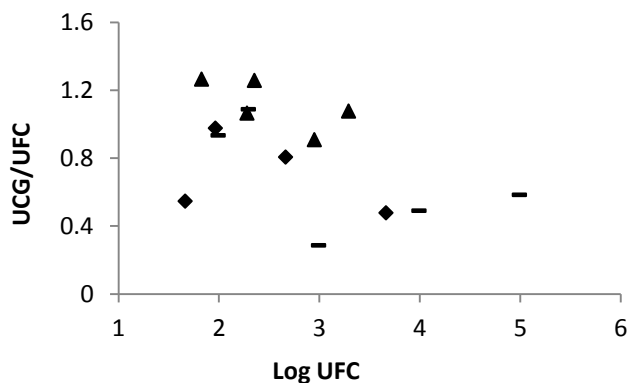


Figure 5.4. Effet de la concentration sur le taux de récupération. *P.aeruginosa* (tiret), *E.coli* (carré), *S.aureus* (triangle)

Deuxièmement, les quelques essais effectués à forte charge bactérienne ont démontré que dépasser les 20 000 UFC, des problèmes d'inhibition de PCR dégrade la qualité du spectre de fluorescence engendrant souvent des résultats erronés. Il est ainsi possible de conclure que la LQ_S se trouvera entre 10 000 et 20 000 UFC. Finalement, la plage de quantification débute à 50 UFC et se termine entre 10 000 et 20 000 UFC.

5.3.6. Robustesse interne

Analyse uni-variée

Les résultats de l'analyse uni-variée se retrouvent dans la Figure 5.5. L'analyse des résultats au niveau de l'impact sur le taux de récupération montre que la sédimentation des cellules dans le contenant de l'échantillon, traitement A vs traitement B, n'est pas un phénomène qui affecte la méthode. L'adsorption des cellules bactériennes à la tuyauterie utilisée pour effectuer la filtration n'affecte pas non plus la méthode. Cependant, lorsque l'on compare le traitement C au traitement D, on constate que la technique de pipetage lors de la re-suspension d'ADN a une influence directe sur les résultats. Le test de Student confirme l'analyse graphique (valeur $p = 0,05$). Les écarts types varient largement d'un traitement à un autre.

Cependant, on ne peut tirer de conclusions par rapport à l'impact du protocole sur la variance étant donné le nombre restreint d'échantillons et sachant que l'écart type relatif moyen, pour les tests de cette version du protocole, oscille autour de 40 à 45 %.

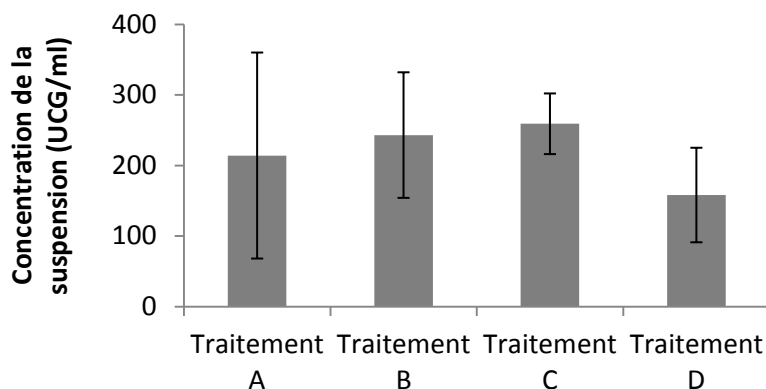


Figure 5.5. Résultats de l'analyse uni-variée.

Analyse avec plan d'expérience factoriel

Les moyennes des niveaux bas (-) et haut (+) de chacun des facteurs étudiés se trouvent dans la Figure 5.6. On constate que les niveaux bas ont tous un meilleur niveau de récupération que les niveaux hauts. Ainsi, en accord avec ces résultats, la présence de bulle dans le filtre, le délai de sonication, le temps de centrifugation et la congélation d'ADN sont des facteurs qui doivent être minimisés pour maximiser le taux de récupération des UCG. L'observation de la Figure 5.6 permet de constater que le temps de centrifugation est le facteur le plus important.

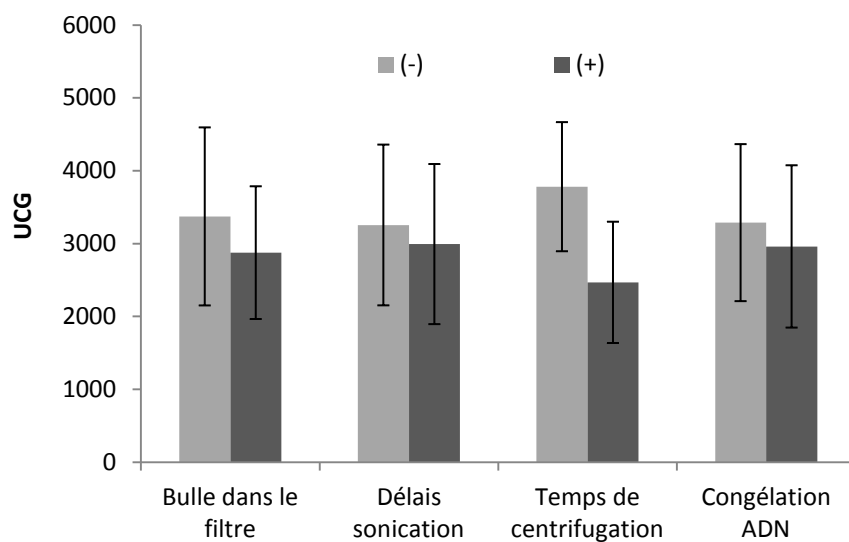


Figure 5.6. Moyenne des niveaux bas (-) et des niveaux hauts (+) des facteurs étudiés pour l'analyse de la robustesse

L'analyse de la variance, Tableau 5.6, montre que l'effet des bulles, le délai de sonication et la congélation d'ADN ne sont pas des facteurs significatifs. Par contre, le temps de centrifugation est un facteur significatif (valeur $p = 0.013$) dont il faut tenir compte.

Tableau 5.6. Plan factoriel 2×2^3 — ANOVA

Analyse de la variance					
Factors	SS	dF	MS	F	Valeur p
Bulle dans le filtre	1,29	1	1,29	1,26	0,29
Délais avant sonication	0,36	1	0,36	0,35	0,57
Temps centrifugation	9,01	1	9,01	8,76	0,01
Congélation ADN	0,56	1	0,56	0,54	0,48
Erreur	11,31	11	1,03		
Total SS	22,52	15			

La grande variance du protocole, 30 à 35 % RSD pour cette version, doit être prise en compte lors de l'interprétation de ces résultats. Ainsi, il est possible que l'utilisation d'un échantillonnage plus grand, qui aurait pour effet de minimiser l'impact de la variance, ait pu permettre de démontrer l'effet significatif d'autres facteurs comme la présence de bulle par exemple. En conséquence et par mesure de précaution, il serait préférable de minimiser la présence de bulle dans le filtre. Aussi, puisque les résultats pour le délai de sonication ont été effectués avec *E.coli* seulement, il est possible que la conclusion ait été différente avec une autre souche puisque la sensibilité de la membrane cellulaire est probablement fonction de l'espèce étudiée. Ainsi, il serait préférable de minimiser l'effet du délai de sonication par précaution, même si l'effet n'est pas significatif pour les conditions étudiées. Finalement la congélation de l'ADN est probablement le seul facteur qui peut être négligé avec certitude. Cependant, l'analyse ne s'applique pas à des cycles de congélations — décongélation qui pourraient avoir un impact différent sur les résultats.

Analyse de données historiques : nettoyage de pièces réutilisables

L'analyse des données historiques des résultats de contrôle de PCR permet d'évaluer l'effet du nettoyage sur une pièce réutilisable en contact avec l'ADN purifié. En comparant les résultats obtenus avant la modification de cette étape de nettoyage contre ceux obtenus après la modification du nettoyage, on est en mesure de vérifier si cette étape est robuste. Les résultats portés en graphique à la Figure 5.7, démontrent visuellement la présence de deux comportements différents incluant une différence de variabilité. Quantitativement, le résultat des contrôles de PCR avant modification du protocole est de 30 ± 30 UCG alors qu'il est de 4 ± 4 UCG après modification. Le test de Student confirme la différence des résultats entre les 2 protocoles (valeur $p = 0.0012$).

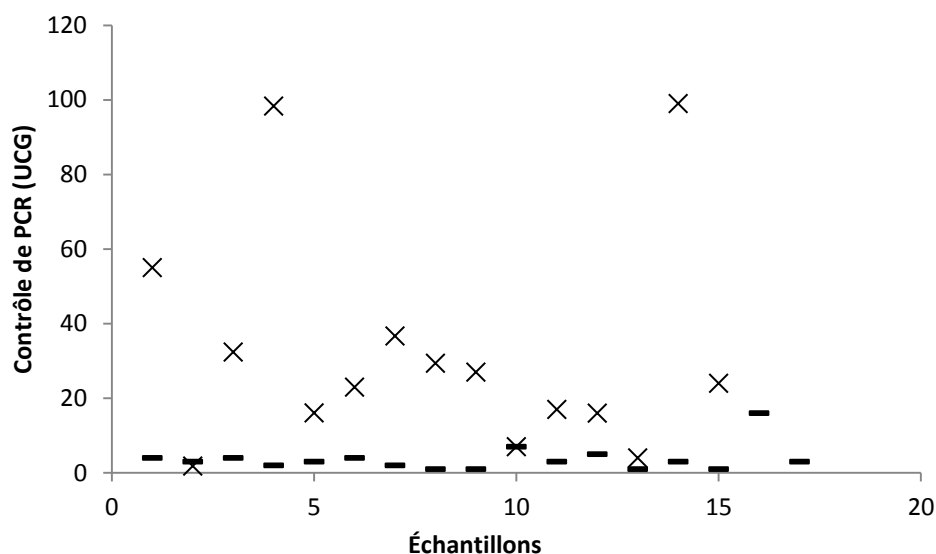


Figure 5.7. Effet du type nettoyage d'une pièce réutilisable sur le résultat des contrôles de PCR. Avant (croix) et après (tiret) la modification du protocole.

La modification du protocole a permis de réduire ce bruit à un niveau non-significatif. Cet ajout de bruit était probablement dû à des contaminations croisées entre expériences de PCR dû à un nettoyage inefficace. Ainsi, l'analyse de données historiques suggère qu'un nettoyage inefficace de la pièce réutilisable augmente le bruit des résultats. L'analyse a permis de démontrer que la modification du protocole a permis d'augmenter la robustesse du protocole et que le nettoyage de cette pièce d'équipement est critique.

5.4. Discussion

L'analyse des performances analytique est largement basée sur une comparaison avec la méthode traditionnelle telle que suggérée par USP. Le taux de récupération utilisé pour évaluer le critère d'exactitude ainsi que le coefficient de détermination utilisé pour évaluer la linéarité souffre du manque de constance du ratio entre les UCG et les UFC. Comme discuté précédemment, section 3.2 et 4.1.3, ce manque de constance est dû à la définition même de ces unités qui sont influencées par des facteurs différents. Néanmoins, les critères d'exactitude et de linéarités sont jugés satisfaisants. Une analyse en profondeur de l'influence de l'état métabolique des cellules, pour une même espèce, sur la variation du ratio UCG/UFC est présentée au chapitre 6 : EFFET DE L'ÉTAT MÉTABOLIQUE. L'analyse dans ce chapitre suggère qu'il serait préférable de comparer les UCG à un compte de cellules viables au lieu d'effectuer une comparaison basée sur les unités formant une colonie.

Le seul critère de performances préoccupantes d'un point de vue analytique est le critère de précision. Un écart type relatif (ETR) inférieur à 35 % constitue l'objectif en terme de précision. Lors des essais initiaux utilisant les premières versions du protocole, la précision (ETR) était de 51% ce qui non conforme aux exigences. Dans la version finale du protocole, la précision calculée, ETR de 33 %, est satisfaisante. Cependant, la valeur calculée est près de la limite acceptable et en considérant l'intervalle de confiance de 6 %, on dépasse cette limite. Une analyse plus approfondie a démontré que 40 % de la variation provenait de l'amplification et de la quantification d'ADN alors que 60 % provenait du protocole d'extraction/purification d'ADN. Dans le chapitre 7 : EFFET DE L'ALGORITHME une analyse des différents algorithmes de quantification permettra d'évaluer si un algorithme différent pourrait améliorer les performances de la quantification et ainsi réduire l'ETR pour augmenter la précision de l'instrument. Au niveau du protocole d'extraction, il est attendu que les variations diminueront lorsque le fournisseur automatisera les étapes d'extraction et de purification. Ainsi, il est attendu qu'une solution commerciale devrait respecter les objectifs de précisions.

L'analyse d'une méthode de qPCR est incomplète si on ne se fit qu'aux critères USP. L'utilisation d'échantillons d'eau stérile enrichie en bactéries viables et cultivables omet l'évaluation de l'impact des bactéries totales sur le résultat de qPCR. Dans le chapitre 9 :

INFLUENCE DES BACTÉRIES TOTALES, l'influence des bactéries totale sur les résultats sera évaluée en utilisant des matières pharmaceutiques. Dans le chapitre 8 : ADAPTABILITÉ DE LA MÉTHODE on présentera comment la méthode de qPCR utilisée a été adaptée afin de faire l'évaluation de matières pharmaceutiques.

5.5. Sommaire

Analyse de performance analytique

- Spécificité — universalité: la méthode a permis de quantifier les huit souches à l'essai respectant ainsi les objectifs de cette évaluation restreinte.
- Exactitude — taux de récupération : toutes les souches testées ont présenté un taux de récupération supérieur à 0.7 UCG/UFC qui était l'objectif qui avait été fixé.
- Précision : l'écart-type relatif entre les répliques du dernier protocole en vigueur est de 33 % avec un intervalle de confiance de 6 %. L'objectif d'être sous les 35% est respecté, mais une amélioration de ce paramètre est souhaitable.
- Linéarité : le R^2 moyen des trois souches testé est de 0.96 ce qui respecte l'objectif d'avoir un coefficient de détermination moyen d'au moins 0.95.
- Plage de quantification : la méthode permet de quantifier adéquatement des échantillons ayant un contenu d'ADN situé entre 50 et 10 000 UCG. Pour les quantités inférieures à 50 UCG, le bruit de fond devient un problème. Pour les quantités supérieures à 10 000 UCG, des problèmes d'inhibition de PCR peuvent survenir.

Éléments en suspens analysés dans les chapitres suivants

- La fluctuation du taux de récupération, pour une même espèce, en lien avec l'état de croissance des cellules affectant le critère d'exactitude et de linéarité sera étudiée plus en profondeur dans le chapitre 6: EFFET DE L'ÉTAT MÉTABOLIQUE.
- L'influence de l'algorithme de quantification sur la précision sera étudiée dans le chapitre 7: EFFET DE L'ALGORITHME
- L'effet des bactéries totales sur les résultats n'était pas l'objet de ce chapitre. Ce phénomène sera étudié plus en détail dans le chapitre 9: INFLUENCE DES BACTÉRIES TOTALES. Le chapitre 8: ADAPTABILITÉ DE LA MÉTHODE permettra d'adapter le protocole pour faire les analyses du chapitre 9.

6. EFFET DE L'ÉTAT MÉTABOLIQUE

6.1. Mise en contexte

Étant donné que les résultats de qPCR sont comparés aux résultats de la méthode traditionnelle, la relation entre les UCG et les UFC doit être comprise afin d'interpréter les résultats adéquatement. Dans la section 5.3.1, les résultats ont montré que la relation entre les UCG et les UFC varient en fonction de l'espèce bactérienne. En plus de varier selon l'espèce bactérienne, il a été supposé que l'âge de la préparation bactérienne avait une incidence sur la relation entre les UCG et les UFC. Ainsi, des essais ont été effectués pour évaluer si l'âge, qui affecte l'état métabolique des cellules, a une influence directe sur cette relation.

6.2. Méthodologie

6.2.1. Terminologie

1. Le terme pente d'agar, ou simplement pente, provient du jargon des microbiologistes et fait référence à un milieu de culture solide coulé dans des tubes de verres de façon à former une pente. Dans le chapitre qui suit, une pente qui n'a pas été inoculé, sera appelé pente vierge ou PV.
2. La culture bactérienne sur une pente fraîche sera définie comme étant une pente qui à été préalablement inoculée avec des bactéries, placée à l'incubateur à 35 °C pendant 16 à 24h, et qui est utilisée durant le jour même où cette pente est sortie de l'incubateur. L'expression «culture sur pente fraîche» portera l'abréviation PF.
3. Une culture sur une pente âgée sera définie comme une pente qui à été préalablement inoculée avec des bactéries, placée à l'incubateur à 35 °C pendant 16 à 24h, puis placée au réfrigérateur plus de 8 h. L'expression «culture sur une pente âgée» portera l'abréviation PA.
4. Une suspension bactérienne provenant d'une pente fraîche ou âgée est obtenue en lavant la PF ou PA avec une solution saline. Une suspension fraîche sera abrégée SF alors qu'une suspension âgée sera appelée SA.
5. En tenant compte des définitions 1 à 4, on comprend que quatre différents types de suspension bactérienne peuvent exister:
 - i. Suspension fraîche provenant d'une culture sur pente fraîche (SF-PF)

- ii. Suspension fraîche provenant d'une culture sur pente âgée (SF-PA)
- iii. Suspension âgée provenant d'une culture sur pente fraîche (SA-PF)
- iv. Suspension âgée provenant d'une culture sur pente âgée (SA-PA)

6.2.2. Essai préliminaire

Trois suspensions bactériennes ont d'abord été analysées pour confirmer si l'âge de la préparation bactérienne avait réellement un impact sur la relation entre les UCG et les UFC. Les suspensions bactériennes ont été créées à partir d'une PF, telle que décrite dans la section 4.2.1, et ont été vieilles de façon à obtenir une suspension de 48 h et une suspension de 8 jours qui seront comparées à une SF. Les différentes préparations bactériennes n'ont pas été répliquées, cependant la méthode d'extraction et d'analyse par qPCR ont été effectuées en triplicata sur chaque pente. Les décomptes sur plaque d'agar ont été effectués en duplicata.

6.2.3. Plan factoriel

Dans le but d'approfondir les essais préliminaires, un plan factoriel a été réalisé pour évaluer l'effet de deux facteurs ayant le potentiel d'affecter l'état métabolique des cellules et ainsi affecter la relation entre les UCG et les UFC. Les deux facteurs étaient l'âge de la suspension bactérienne et l'âge de la culture sur pente d'agar. Les deux facteurs ont été variés sur deux niveaux. Étant donné que les analyses de qPCR ont une variance élevée, les traitements ont été testés en quadruplicata. Les cultures sur pentes n'ont pas été répliquées, cependant la méthode d'extraction et d'analyse par qPCR a été effectuée en triplicata sur chaque pente. Les tests sur plaque d'agar ont été effectués en duplicata. Les résultats ont été analysés en se basant sur le ratio entre les UCG et les UFC. Une transformation des données de type box-cox a été nécessaire pour stabiliser la variance avant de faire l'analyse ANOVA.

Pour des questions pratiques, la moitié des échantillons d'ADN purifié a été congelée et analysé le lendemain contrairement à la première moitié qui fut analysée le jour même. Puisque les échantillons sont séparés en deux groupes, on se trouve à couper la moitié des répliques pour créer un facteur bloc. Ce facteur bloc sera représentatif de l'effet combiné de la congélation et de l'amplification d'ADN sur deux séances différentes. Ainsi, les 16 échantillons ont été séparés en deux blocs. Les traitements sont résumés dans le tableau 6.1.

Tableau 6.1. Plan 2x3² : traitements

Traitement	Nomenclature	Âge de la suspension (Heures)	Âge de la pente (Jours)	Congélation (Bloc)
Cas A	SF-PF	0	0	NON
Cas B	SA-PF	48	0	NON
Cas C	SF-PA	0	7	NON
Cas D	SA-PA	48	7	NON
Cas A	SF-PF	0	0	OUI
Cas B	SA-PF	48	0	OUI
Cas C	SF-PA	0	7	OUI
Cas D	SA-PA	48	7	OUI

6.3. Résultats

6.3.1. Essai préliminaire

Les résultats préliminaires, Tableau 6.2, montrent que le nombre d'UCG diminue plus rapidement avec l'âge de la suspension qu'avec le nombre d'UFC. Ainsi, le rapport UCG/UFC est très large pour une SF et diminue rapidement avec l'âge de la suspension. Cet essai préliminaire confirme l'influence de l'état de croissance des cellules sur la relation entre les UCG et les UFC et justifie une analyse plus poussée en plan factoriel.

Tableau 6.2. Influence de l'âge de la suspension bactérienne

Âge de la suspension	qPCR (UCG) (n=3)	Décompte (UFC) (n=2)	UCG/UFC
8 jours	1 700	260	6,8
24 h	3 800	290	13
Fraiche	8 700	170	53

6.3.2. Plan factoriel

Tel que décrit dans la section 6.2.3, un plan factoriel 3^2 en duplicats a été réalisé. Comme montré dans les essais préliminaires, le nombre d'UCG augmente avec l'âge de la suspension pour une suspension provenant d'une PF. Cependant, pour une PA, l'inverse se produit, le nombre d'UCG augmente avec l'âge de la suspension. Si l'on observe l'évolution du ratio UCG/UFC, on constate la même tendance. Ainsi, on peut supposer un effet d'interaction entre l'âge de la culture sur pente et l'âge de la suspension bactérienne (Figure 6.1).

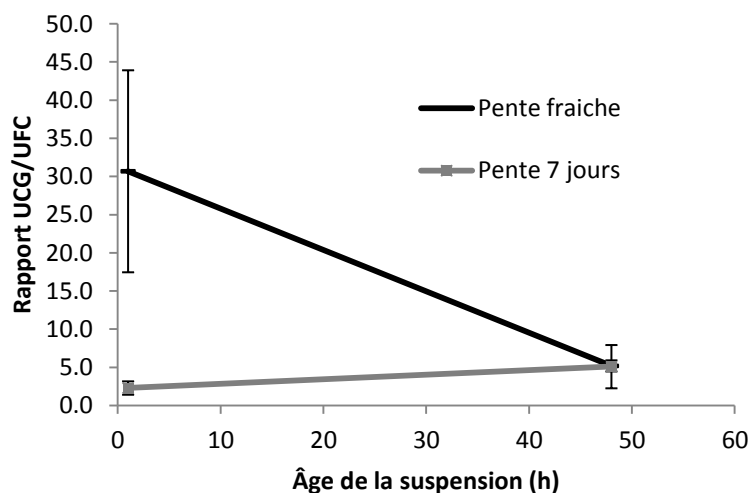


Figure 6.1. Interaction entre l'âge de la culture sur pente et l'âge de la suspension bactérienne

L'analyse de la variance, Tableau 6.3, permet d'évaluer avec précision l'effet des interactions et de vérifier quels facteurs sont significatifs. L'analyse montre que l'effet du bloc (effet de la congélation avant analyse) est non significatif alors que l'effet de l'âge de la culture sur pente ainsi que l'effet de l'âge de la suspension sont significatifs (valeur $p = 0.0278$ et 0.000213 respectivement). L'effet de l'interaction entre l'âge de la suspension et l'âge de la culture de la pente serait aussi significatif que l'effet de l'âge de la suspension (valeur $p = 0.000606$). On conclut que pour minimiser le rapport entre les UCG et les UFC, l'âge de la suspension doit être minimisé en maximisant l'âge de la culture sur pente.

Tableau 6.3. Plan factoriel 2x2³ 16 essais — ANOVA

Analyse de la variance					
Facteurs	SS	dF	MS	F	Valeur p
(1) Congélation (Bloc)	11.5	1	11.5	1.15	0.307000
(2) Âge de la pente	64.2	1	64.2	6.42	0.027800
(3) Âge de la suspension	293	1	293	29.3	0.000213
(2-3) Interaction	225	1	225	22.5	0.000606
Erreur	110	11	10.0		
Total SS	704	15			

6.4. Discussion

6.4.1. Comportements observés

Cette expérience a permis de mettre en lumière les phénomènes suivants:

1. Pour une PF, le rapport UCG/UFC diminue avec l'âge de la suspension qui a été créé à partir de la pente (SF-PF vs SA-PF).
2. Pour une PA, le rapport UCG/UFC augmente avec l'âge de la suspension (SF-PA vs SA-PA).
3. L'analyse ANOVA démontre une forte interaction entre l'âge de la culture et l'âge de la suspension.
4. Le rapport UCG/UFC est à son maximum pour une suspension fraîche provenant d'une culture fraîche sur pente d'agar (SF-PF).
5. Le rapport UCG/UFC est à son minimum pour une suspension fraîche provenant d'une vieille culture (SF-PF).

Pour interpréter ces résultats, il faut tenir compte des facteurs suivants :

- I. Les bactéries provenant d'une nouvelle culture sur pente d'agar sont en phase exponentielle de croissance alors que les bactéries d'une vieille culture sont en phase stationnaire.
- II. Le nombre d'UFC compté est inférieur au nombre de cellule cultivable lorsqu'il y a agrégation cellulaire (39-41).
- III. Le nombre d'UCG tient compte des cellules totales.

6.4.2. Hypothèses retenues

La prise en considération des facteurs I à III permet de poser des hypothèses dans le but de tenter d'expliquer les phénomènes 1 à 5. En considérant que les bactéries se trouvent dans des agrégats cellulaires, en considérant que le nombre de cellules vivantes par agrégat cellulaire est dépendant de l'état de croissance des cellules et en considérant les différences fondamentales entre les UCG et les UFC, on arrive à formuler une hypothèse qui tente d'expliquer les comportements observés.

Explication de l'effet de l'âge de suspension

Supposons que pour une SF-PF, on trouve 100 agrégats de cellules par millilitre, que chaque agrégat contient 10 cellules et que chaque cellule contienne 1UCG. On obtiendra un décompte de 100 UFC/ml ainsi qu'une quantification par qPCR de 1000 UCG/ml. Supposons maintenant une SF-PA. Dans ce cas, on doit aussi tenir compte de la mortalité, ainsi que de la lyse cellulaire. Ainsi, même si le nombre d'agrégats est relativement constant, le nombre de cellules par agrégat aurait diminué considérablement dû à la lyse cellulaire. Donc, le rapport UCG/UFC aura diminué pour une SA par rapport une SF.

Explication de l'effet de l'âge de la culture sur pente

Considérons maintenant le cas d'une suspension provenant d'une SF-PA. Comme dans le cas d'une SA-PF, la mortalité cellulaire ainsi que de la lyse cellulaire est à prévoir. Ainsi le nombre de cellules par agrégats devrait diminuer. Donc le rapport UCG/UFC diminuerait par rapport à la suspension de départ.

Explication de l'interaction entre l'âge de la suspension et l'âge de la culture sur pente

Dans le cas d'une SA-PA, on supposera que les agrégats, provenant de la PA, qui sont mis en suspension contiennent peu de cellules. Cependant, le fait de gratter la pente d'agar pour détacher les cellules et de re-suspendre le film bactérien dans une saline change les conditions de cultures. Ainsi, les agrégats en suspension se retrouvent dans un milieu appauvri en déchets cellulaires, dus à l'effet de dilution, et enrichi en nutriments dû au grattage de la gélose. Donc, il est possible que le nombre de cellules par agrégat cellulaire augmente dû à la croissance cellulaire. La Figure 6.2 illustre et résume les explications précédentes.

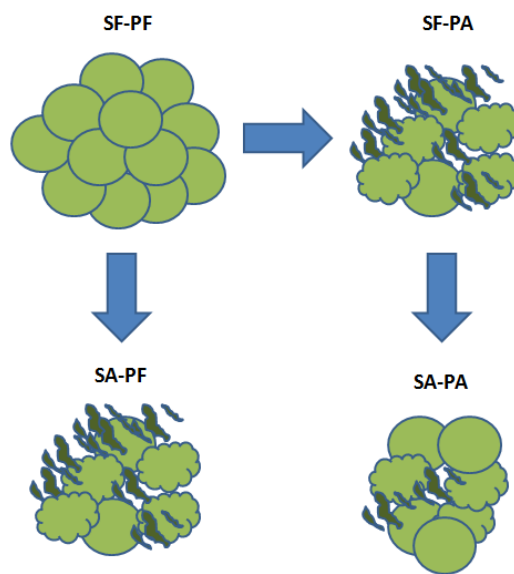


Figure 6.2. Effet hypothétique de l'âge de la culture et de la suspension sur les agrégats cellulaires

L'hypothèse de l'agrégat à niveau de viabilité variable pourrait être confirmée par des analyses de microscopie avec un marquage de type *live/dead*. Parmi les hypothèses soulevées, celle-ci pourrait être la plus probable selon l'auteur.

6.4.3. Autres hypothèses

Hypothèse de division cellulaire durant l'exécution du protocole

On pourrait supposer que la division cellulaire ne doit pas être négligée durant l'intervalle de temps débutant de l'étape où les bactéries sont mises sur plaque d'agar jusqu'à l'étape où l'ADN des bactéries est extrait pour les tests de qPCR. Cependant, aucun lien n'a été décelé entre l'ordre des échantillons et le résultat de qPCR, Figure 6.3. Donc cette hypothèse est rejetée. Notez que l'ordre expérimental des échantillons était aléatoire face au traitement du plan d'expérimentation.

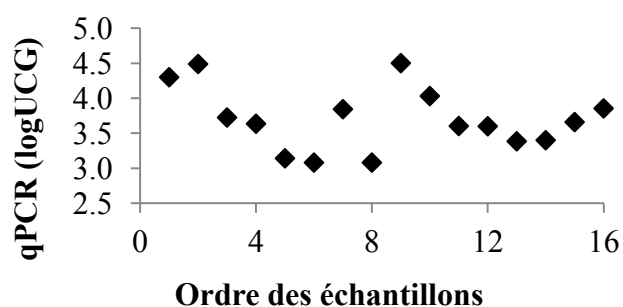


Figure 6.3. Effet de l'ordre des échantillons sur le résultat final

Hypothèse de la variabilité du phénomène d'agrégation cellulaire

On pourrait aussi supposer que l'agrégation cellulaire varie selon l'état de croissance des cellules, ce qui expliquerait les changements du ratio UCG/UFC. Cependant, puisque le décompte est relativement constant entre les quatre cas de figure (RSD = 12 %) et que la quantification par qPCR varie fortement entre les cas de figure (RSD=114%), cette hypothèse est rejetée.

6.5. Sommaire

Les résultats de ce chapitre ont permis d'apprendre que l'âge de la préparation bactérienne avait un impact significatif sur la divergence entre les UCG et les UFC. En effet, l'âge de la culture bactérienne ainsi que l'effet de l'interaction entre l'âge de la culture bactérienne en milieu solide et l'âge de la suspension dans une solution saline ont un effet très significatif sur le ratio UCG/UFC.

L'hypothèse la plus plausible expliquant le lien entre l'âge et le ratio UCG/UFC est la présence d'agrégats de bactéries dont le nombre de cellules varie en fonction des conditions de milieux qui sont favorables ou défavorables à la croissance cellulaire. Ainsi un ratio UCG/UFC élevé signifie que le nombre de bactéries par agrégat est élevé et que le métabolisme cellulaire est en phase de croissance. À l'inverse, un ratio plus faible montre une population bactérienne en phase stationnaire ou de déclin.

Ces conclusions sont importantes pour ceux qui feront des tests de validation avec le qPCR. Lors des essais de validation, il serait important de maintenir un ratio UCG/UFC constant pour une même espèce. Ainsi un protocole de préparation des cellules devra être monté pour minimiser les fluctuations de ce ratio. Une façon de faire pourrait être d'utiliser des bactéries en phase de latence, ou encore d'avoir un protocole minimisant les agrégats. D'après les informations de notre fournisseur, les bactéries commercialement vendues sous forme de billes, *Bioballs*™, permettent d'obtenir un ratio UCG/UFC constant. Pour ce qui est des travaux présentés, le protocole fut modifié afin d'utiliser des suspensions bactériennes provenant de culture sur pente qui ont été placées au réfrigérateur pendant 7 jours après avoir été sorties de l'incubateur (SF-PA).

7. EFFET DE L'ALGORITHME

7.1. Mise en contexte

Lorsque la réaction de PCR est terminée, les données brutes de fluorescences sont analysées par un algorithme mathématique qui détermine le cycle de quantification (C_q) de celle-ci. En utilisant les valeurs de la courbe standard, le C_q de chaque réaction de PCR est converti en nombre d'UCG. La robustesse de l'algorithme de quantification a une influence directe sur la précision, le biais et l'erreur des résultats. Selon l'analyse de la précision, section 5.3.3, 40 % de la variance des répliques serait attribuable à l'algorithme de quantification. Ainsi, les algorithmes de quantification de PCR ont été étudiés afin d'évaluer si un algorithme supérieur pouvait remplacer celui utilisé actuellement en vue d'améliorer les résultats de la méthode.

Des algorithmes existant dans la littérature [Tichopad *et al*, 2003; Wilhelm *et al*, 2003b; Guescini *et al*, 2008; Wilhelm *et al*, 2003b] et quelques algorithmes développés spécifiquement pour la méthode ont été évalués. Au final, un algorithme basé sur la détermination du C_q par lissage et différentiation de Savitzky-Golay (LDSG) a été développé et semble démontrer un potentiel intéressant pour l'analyse des données brutes de fluorescence. L'algorithme développé a été comparé, au niveau de ses performances, à trois différentes méthodes de quantification connues dans la littérature ainsi qu'à la méthode du fournisseur.

7.2. Méthodologie

7.2.1. Développement et évaluation comparative de l'algorithme

Brièvement, l'algorithme LDSG consiste à utiliser l'abscisse du maximum de la dérivée, obtenue par lissage et différentiation de Savitzky-Golay, comme C_q de la réaction de PCR. Les données brutes sont préalablement augmentées par interpolation linéaire afin d'augmenter le nombre de décimales du C_q . L'algorithme a été optimisé en utilisant deux plans factoriels qui ont permis d'évaluer quels paramètres doivent être utilisés pour maximiser la précision et minimiser l'erreur et le biais. Les paramètres évalués sont l'ordre de la dérivée, l'ordre du polynôme de lissage, la largeur de la fenêtre de lissage et le nombre de points interpolés entre les points expérimentaux pour l'ajout de décimale au C_q .

L'algorithme LDSG a été comparé à la méthode du *threshold* (C_t), à la méthode du fournisseur de l'appareil, ainsi qu'à deux méthodes basées sur la modélisation de la courbe sigmoïde : une méthode de modélisation de la courbe sigmoïde à quatre paramètres (MS4P) et une méthode basée sur l'intersection entre le point d'inflexion et l'abscisse (C_{y0}). Les différents algorithmes ont été évalués par rapport à l'erreur, le biais, la précision, la normalité des résidus et l'effet des *competimers*. Les données utilisées pour évaluer l'effet de l'algorithme ont été sélectionnées de façon à maximiser les variations inter-essais.

La méthodologie et les résultats concernant le développement de l'algorithme LDSG ainsi que la comparaison avec d'autres algorithmes de la littérature font l'objet d'un article scientifique. Conséquemment, la méthodologie détaillée ainsi que les résultats de ce chapitre se retrouvent dans l'article qui est inséré dans ce chapitre. Étant donné le caractère confidentiel de cette étude, la comparaison avec l'algorithme du fournisseur n'a pas été incluse dans l'article, cependant l'effet du nouvel algorithme sur les résultats de la méthode de qPCR sera discuté dans la section 7.5 qui suivra l'article.

7.2.2. Évaluation de l'impact sur la précision de la méthode de qPCR

Dans le but d'évaluer l'effet du nouvel algorithme sur la précision de la méthode de qPCR, les données brutes de fluorescence, provenant des tests d'échantillons contenant des bactéries, ont été ré-analysées avec le nouvel algorithme pour faire une comparaison avec les résultats de l'algorithme de référence. La comparaison évaluera la précision de la méthode pour le protocole 1 qui était sujet à des problèmes d'inhibition, pour le protocole 2 qui n'était pas optimal et pour le protocole 3 qui est le protocole finale.

Les données provenant du protocole #1 sont les cinq suspensions en triplicata de *S. aureus* et de *P. aeruginosa* qui ont été utilisés pour l'évaluation de la linéarité ($n=30$). De ces données, les résultats de treize échantillons avec inhibition de PCR n'ont pu être utilisés. Ainsi, l'analyse pour ce protocole comporte 23 échantillons. Les données provenant du protocole 2 contiennent tous les échantillons de 2012 en excluant ceux du protocole #1. Ainsi, l'analyse du protocole 2 est réalisée sur 104 échantillons. Les données du protocole 3 utilisent 52 échantillons de tests effectués en 2013 et excluent les essais sur des matières pharmaceutiques.

Puisque l'analyse a été effectuée avec des données de fluorescence déjà existante, le lecteur peut consulter la section 4.2 pour s'informer sur la procédure générale. Pour plus d'information au niveau des souches, voir les sections 5.2 et 6.2. Pour un sommaire des différences entre le protocole 1, 2 et 3, voir l'ANNEXE B.

7.3. Résultats

7.3.1. Développement et évaluation comparative de l'algorithme

Auteurs et affiliation :

Charles GAUDREAULT, Étudiant à la maîtrise en génie chimique, Université de Sherbrooke, Faculté de Génie, Département de génie chimique et biotechnologique, 2500 boul. de l'université, Sherbrooke, QUÉBEC, CANADA J1K 2R1

Joanny SALVAS, scientifique sénior, Process Analytical Sciences Group, Pfizer Montréal, 1025 boul. Marcel-Laurin, Montréal, QUÉBEC, CANADA, H4R 1J6

Joël SIROIS, Professeur agrégé, Université de Sherbrooke, Faculté de Génie, Département de génie chimique et biotechnologique, 2500 boul. de l'université, Sherbrooke, QUÉBEC, CANADA J1K 2R1

Date de soumission : 2015-12-07

État de l'acceptation : Article soumis

Revue : BMC Bioinformatics

Titre français : Lissage et différentiation de Savitzky-Golay pour la quantification de données de PCR

Résumé français:

Contexte: Le recours au PCR quantitatif (qPCR) à des fins d'analyses a augmenté rapidement au cours de la dernière décennie. Un manque d'amélioration au niveau de l'exactitude et de la précision pourrait cependant freiner l'augmentation de son utilisation. La quantification à partir de données de fluorescence est affectée par des variations inter et intra essais. L'utilisation de répliques permet de minimiser l'impact des variations intra-essais, cependant les variations inter essais doivent être minimisées afin d'obtenir une méthode de quantification robuste. La méthode proposée dans cette étude utilise le lissage et la différentiation de Savitzky-Golay (LDSG) pour identifier un cycle de quantification basé sur le maximum d'une dérivée. La méthode ne dépend pas d'une modélisation de la courbe de fluorescence, comme c'est généralement le cas pour plusieurs méthodes existantes.

Résultats: Des séries de données de fluorescence de PCR soumises à plusieurs variations inter essais (différents thermocycleurs, différents lots de réactifs, différents opérateurs, différentes courbes standard et différents laboratoires) ont été utilisées pour l'évaluation. En premier lieu, l'erreur relative (ER), le biais relatif (BR) et l'écart type relatif (ETR) de l'algorithme de la méthode LDSG proposé a été minimisé en utilisant un plan d'expérience (DoE). Le DoE a été utilisé pour évaluer l'effet de quatre paramètres : ajout de point de donnée entre les données expérimentales en utilisant une interpolation cubique, largeur de la fenêtre de lissage, ordre du polynôme de lissage et ordre de différentiation. Les paramètres identifiés permettant de minimiser ER, BR et ETR sont: addition de 7 points de données entre les données expérimentales, utilisation d'un polynôme d'ordre 1 et différentiation de premier ordre. La largeur optimale de la fenêtre de lissage ne peut être généralisée. En second lieu, l'algorithme a été comparé à la méthode du modèle logistique à quatre paramètres (4PLM), à la méthode du C_{y0} et à la méthode du *threshold*. La méthode LDSG se compare avantageusement aux autres méthodes lorsque soumise à des variations inter-essai. L'ER est statistiquement différent de 4PLM (valeur-p = 0.03), de C_{y0} (valeur-p = 0.05) et de la méthode du *threshold* (valeur-p = 0.004). Lorsque soumis à des variations intra essais, LDSG est supérieur à la méthode du *threshold* (valeur-p = 0.005) mais est équivalent à 4PLM et C_{y0} (valeur-p > 0.05).

Conclusion: Nos résultats démontrent que les performances de la méthode LDSG dépasse de façon significative celle des méthodes basées sur la modélisation de la courbe sigmoïde (4PLM et C_{y0}) et de la méthode du *threshold* lorsque les données de PCR sont affectées par des variations inter essais. Ainsi, la méthode devrait être recommandée pour minimiser les variations inter essais pour les protocoles de qPCR sujet à ce type de variation.

Savitzky-Golay smoothing and differentiation for PCR quantification

Charles Gaudreault¹, Joanny Salvas² and Joël Sirois¹

Address: ¹Engineering faculty, Université de Sherbrooke, 2500 boul. de l'université, QUÉBEC, CANADA J1K2R1, ²Process Analytical Science Group, Pfizer Montréal, 1025 boul. Marcel-Laurin, Montréal, QUÉBEC, CANADA, H4R1J6

Corresponding author: Joel.Sirois@USherbrooke.ca

Keyword : gene quantification, PCR quantification, quantification cycle, qPCR, Savitzky-Golay, polymerase chain reaction, real-time PCR, absolute quantification, smoothing and differentiation

Abstract

Background: The use of quantitative PCR (qPCR) analysis has rapidly increased in the past decade. Lack of improvement in accuracy and precision could hinder the spreading of its use. Quantification from fluorescence data is affected by intra- and inter-assay variations. Replicates can minimize the impact of intra-assay variation, however inter-assay variations must be minimized in order to obtain a robust quantification method. The method proposed in this study uses Savitzky-Golay Smoothing and Differentiation (SGSD) to identify a derivative maximum-based cycle of quantification. It does not rely on curve modeling, as is the case with many existing techniques.

Results: PCR fluorescence data sets challenged for inter-assay variations (different thermocycler units, different reagents batches, different operators, different standard curves and different labs) were used for the evaluation. As a first step, the proposed SGSD method algorithm's relative error (RE), bias (RB) and standard deviation (RSD) were minimized using a design of experiment (DoE). The DoE was used to evaluate the effect of four parameters : addition of data points between experimental points using cubic interpolation, smoothing width, smoothing polynomial order and differentiation order. The parameters identified to minimize RE, RB and RSD were : addition of seven data points by interpolation and use of first order smoothing polynomial and first order differentiation. The smoothing width was

found to be specific to the assay. As a second step, the algorithm was then compared to a four-parameters logistic model method (4PLM), the C_y0 method and the threshold method. The SGSD method compared favorably to all methods in terms of inter-assay variation. RE was statistically different from 4PLM (p-value = 0.03) C_y0 (p-value = 0.05) and threshold method (p-value = 0.004). For intra-assay variations, RE outperformed the threshold method (p-value = 0.005) and equalled the 4PLM and C_y0 methods (p-value > 0.05).

Conclusions : Our results demonstrates that the SGSD method performs significantly better than sigmoid modeling based methods (4PLM and C_y0) and threshold method when PCR data are challenged for inter-assay variations. Thus, the method should be recommended to minimize inter-assay variation from assay prone to suffer from it.

Background

Since its development by Kary Mullis in 1983 [Mullis, 1993], DNA amplification by polymerase chain reaction (PCR) and its variant, quantitative PCR (qPCR), have been applied in diverse fields related to molecular biology [Dolan *et al.*, 2009; Kephart et Bushon 2009; Jimenez *et al.*, 2000; Martinon *et al.*, 2012; Hierro *et al.*, 2007; Sivaganensan *et al.*, 2012; Pathak *et al.*, 2012; Haugland *et al.*, 2005; Dusserre *et al.*, 2008; Walker *et al.*, 1998]. Quantitative PCR, which appeared in the 90' [Higuchi *et al.*, 1992; Higuchi *et al.*, 1993], is based on real-time fluorescence measurement of DNA template amplification. Quantification is performed using mathematical treatment of the raw fluorescence data. High sensitivity, theoretically as low as 3 gene copies [Birch *et al.*, 2001], and high throughput, have made qPCR the gold standard in terms of DNA quantification in different fields [Dolan *et al.*, 2009; Kephart et Bushon, 2009; Jimenez *et al.*, 2000; Espy *et al.*, 2006; Postollec *et al.*, 2011; Jimenez, 2011].

Although very sensitive, qPCR is also prone to error from two major sources. The first source of error results from efficiency determination which is used to calculate the output quantity. Fluctuation in the amplification curve shape affects efficiency calculations as well as the quantification process [Sisti *et al.*, 2010]. This efficiency can be calculated from a standard curve plot [Wilhelm et Pingoud, 2003] as usually performed for absolute quantification. Another way is to calculate it individually by numerical analysis of the reaction exponential

phase as performed for relative quantification [Tichopad *et al.*, 2003] and sometime for absolute quantification [Rutledge, 2004]. Relationship between the efficiency (E), the DNA template concentration (T) and the number of amplification cycle (n) is given by the following equation:

$$T_n = T_0(1 + E)^n \quad (1)$$

It is understood that any difference between the calculated efficiency and the real efficiency would have an important effect on the calculated DNA template concentration. This effect becomes more important with higher number of amplification cycles. Standard curves are used under the assumption that efficiency is constant between replicates, while numerical efficiency calculation takes these fluctuations into account.

The second major source of error comes from the shape of the amplification curve subject to experimental variation affecting gene quantification. This is generally true whether quantification is calculated from a standard curve (absolute quantification) or based on a housekeeping gene comparison (relative quantification). The amplification curve behaves like a logistic curve and is subject to experimental variations in asymmetry [Spiess *et al.*, 2008], slope steepness at inflexion, location of inflexion point and total increase in fluorescence from baseline to plateau [Sisti *et al.*, 2010; Guescini *et al.*, 2013]. To eliminate the curve comparison bias, it has been shown that a quantification methods relying on optical calibration rather than on curve comparison can be developed [Rutledge, 2004].

In the first decade of the 21st century, efforts were made to increase qPCR precision. Breakthrough in relative quantification paved the way for quantification without standard curves [Pfaffl, 2001]. This innovation prompted the need to correct the baseline [Wilhelm *et al.*, 2003a] and to model the logistic curve so as to determine the PCR efficiency directly from the amplification curve [Tichopad *et al.*, 2003; Liu et Saint, 2002]. Meanwhile, absolute quantification, based on standard curves using the threshold method [Wilhelm et Pingoud, 2003; Rutledge et Côté, 2003] was adapted for logistic curve fitting [Goll *et al.*, 2006] and made more robust against PCR inhibitors [Guescini *et al.*, 2008]. Numerous authors have worked to increase precision of both quantification methods [Sisti *et al.*, 2010; Spiess *et al.*, 2008; Gescini *et al.*, 2013; Ruijter *et al.*, 2013; Carr et Moore, 2012; Karlen *et al.*, 2007]. Over

the past years, many different approaches were used for quantification of PCR data [Ruijter *et al.*, 2013] and new methods are still being developed and tested [Rutledge, 2004; Carr et Moore, 2012; Franke *et al.*, 2012] since qPCR data variance is not yet fully explained.

The approach described in this article investigates a non-modeling avenue for quantification. This contrasts with most quantification methods, which are based on PCR curve modeling. In this study the PCR curve and its derivatives were considered to behave like spectral bands with a certain level of noise. Thus, the Savitzky-Golay smoothing and differentiation (SGSD) filter commonly used for spectral band noise removal [Savitzky et Golay, 1964; Schafer, 2011] was evaluated as a mean to determine a derivative maximum-based cycle of quantification. To the best of our knowledge, this avenue has been scarcely researched and never thoroughly evaluated [Franke *et al.*, 2012; Trampuz *et al.*, 2006]. The parameters of the SGSD filter were evaluated and optimized, and the method was then compared to three other PCR quantification methods: the threshold method, the four-parameters logistic curve model method, and the C_{y0} method. The investigated methods are compared from an absolute quantification perspective using several criteria (relative bias, error and standard deviation), the effect on outlier and the effect of competitors

Methods

Fluorescence data

Data were obtained in two steps. As a first step, series of data were obtained from three different labs using the same equipment and method. Each PCR run consists of duplicated serial dilutions of standard DNA at 10, 50, 100, 1 000 and 10 000 gene copies (G_c). A total of 20 runs, called data set in this article, were used, for a total of 200 PCR amplification curves as shown by figure 1. In a second step, the quantification methods were challenged with competitor and different PCR assays using data sets from the literature.

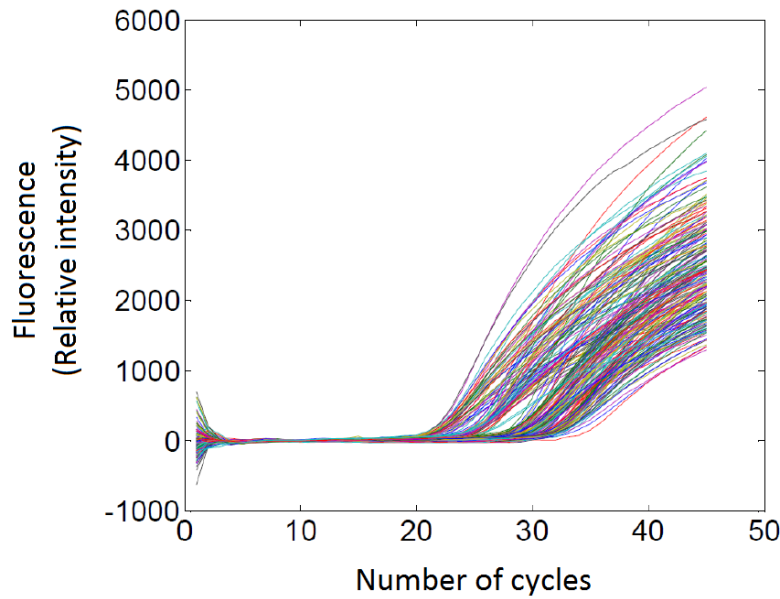


Figure 1. The 200 raw fluorescence PCR curves, with baseline adjusted to zero, used to evaluate quantification algorithms.

In order to validate results, three data sets were obtained from literature. The data set from Jan M. Ruijter [Ruijter *et al.*, 2013] consists of MYCN gene of 10-fold dilutions from 15 000 to 15 G_c for 94 replicates. Data sets from Bob Rutledge [Rutledge, 2004] consist of 10-fold dilutions of 4.17×10^7 to 4.17×10^2 copies in quadruplicate for 5 runs of the K1/K2 and K2/K3 amplicons. Dilution 4.17×10^7 was excluded since it generated C_q too low to allow fair comparison against the threshold method.

In order to simulate decreased efficiency caused by inhibitors, competitive primers (competimers) have been used by previous authors [Ruijter *et al.*, 2013] to change PCR

efficiency in a predictable way. A competitor data set was obtained from the literature [<http://qPCRDataMethods.hfr.nl>]. The preparation of this data set has been described in Jan M. Ruitjer. The data set contained results with 0.0625 µg DNA by 4-fold dilution in triplicate for competitor concentration covering 0 to 50 %.

Data pre-treatment

Raw fluorescence data were screened using MATLAB to identify lack of PCR amplification. To do so, the mean fluorescence value of baseline (n_3 to n_7) was compared to the mean value of plateau (n_{40} to n_{45}). The plateau had to be 10 standard deviations (SD) higher than the baseline for the curves to be included into the datasets. The baseline of the curves obtained using the 4PLM and the C_y0 methods were adjusted to zero before data processing. The baseline was adjusted with a saturation function such as described by Wilhelm [Wilhelm *et al.*, 2003b]:

$$F_{(n)0} = a(1 - e^{b-n}) + c \quad (2)$$

where $F_{(n)0}$ is the baseline at cycle n , a the saturation value, b the number of cycle for baseline saturation, c the fluorescence at saturation. The equation was fitted to the curve using non-linear least-square fitting by varying a , b and c . Additional file 1 describes the baseline adjustment using a MATLAB script.

Standard curves for calculations of number of gene copies

Standard curves are made with the fluorescence data in order to correlate calculated C_q with the number of gene copies (G_c) originally present in the sample. Usually, a standard curve is prepared once using one data set of ten standard DNA sample and it is used to estimate unknown DNA sample. Using a single standard curve for our study could potentially favor one quantification method over another. In order to minimize the probability of introducing a bias in the study, ten data sets were used out of 20 data sets to generate ten standard curves. One by one, each one of these ten data sets were used to estimate the DNA content of the other 19 data sets (190 fluorescence curves). At the end of the process, 10 estimates were created for each DNA sample. The ten estimates of the DNA content in an experiment were then averaged. This data processing method minimizes the possible bias introduced by a single standard curve. Each calibration curve was done with amplification curves obtained with 10, 50, 100, 1 000 and 10 000 G_c from the same sample preparation. Calibration curves were

generated by linear regression of the logarithm of the standard DNA G_c against the output C_q . The linear regression equation is given by equation (3) which can be rearranged into the more practical equation (4).

$$C_q = m * \text{Log}_{10}G_c + b \quad (3)$$

$$G_c = 10^{\frac{(C_q-b)}{m}} \quad (4)$$

where m and b are respectively the slope and the intercept of the standard curve.

Description of 4PLM method

The algorithm consists of a sigmoid curve fitting to the amplification curve as described by Ales Tichopad [Tichopad *et al.*, 2003]. PCR amplification curves were fitted using the following equation:

$$F_{(n)} = F_0 + \frac{\Delta F}{1 + \left(\frac{n}{n_0}\right)^{-b}} \quad (5)$$

where F is the fluorescence, n the cycle number, F_0 the baseline fluorescence, ΔF the difference between the plateau fluorescence and the baseline, n_0 the inflexion point and b describes the slope at x_0 . The equation was fitted using MATLAB non-linear least-square fitting and the default algorithm called *Thrust Region Effective* [Coleman et Li, 1996]. The initial estimates for iteration are : $F_0 = \min(F_{(n)})$, $\Delta F = \max(F_{(n)}) - \min(F_{(n)})$, $b = 1$ and the estimate for n_0 was given by n at $F_{(n)}$ given by equation (6).

$$F_{(n_0)} = \min \left\| F_{(n)} - \frac{[\max(F_{(n)}) - \min(F_{(n)})]}{2} \right\| \quad (6)$$

The algorithm's lower and upper bound were set to:

$$\begin{aligned} F_0 - \Delta F &< F_0 < F_0 + \Delta F \\ \Delta F - 0.01 * \Delta F &< \Delta F < \Delta F + 0.01 * \Delta F \\ 0 &< b < \infty \\ 0 &< x_0 < \infty \end{aligned}$$

The method used for this article differs from Tichopad's as C_q was set here to the maximum of the 2nd order derivative. This was realized by first generating new data points using the fitted equation. Then the data points were differentiated twice using the MATLAB *differentiate* function. The second derivative maximum was found using the MATLAB *max* function on

these differentiated data points. Additional file 2 describes the 4PLM method MATLAB script.

Description of C_y0 method

The C_y0 method has been described in detail elsewhere [Guescini *et al.*, 2008]. The PCR curve is first fitted to a five-parameter logistic model given by the Richard's equation:

$$F_{(n)} = F_0 + \frac{\Delta F}{\left[1 + e^{\left(\frac{-1}{b}(n - n_0)\right)}\right]^d} \quad (7)$$

where F is the fluorescence, n the cycle number, F_0 the baseline fluorescence, ΔF the difference between the plateau fluorescence and the baseline fluorescence, n_0 the inflexion point, b describes the slope at n_0 and d is the Richard's coefficient. As for the 4PLM, the equation was fitted using MATLAB non-linear least-square fitting with initiation estimates and bound parameters. The initial estimates and the bounds were the same as for the 4PLM while initial estimate for d was 1 bounded as $0 < d < 10$. The cycle of quantification, C_y0, is the intercept where the tangent of the inflexion point crosses the x-axis with a baseline set to zero. The C_y0 is given by equation (8). Additional file 3 describes the C_y0 method MATLAB script.

$$C_{y0} = n_0 - b \left(1 + \frac{1}{d} - \ln d\right) \quad (8)$$

Description of the threshold method (C_t)

In this method, the cycle of quantification is defined to be the cycle at which the fluorescence reaches an established threshold. The baseline's (n_3 to n_{15}) standard deviation (SD) of every amplification curve was calculated and used to establish a threshold set at 10SD. Then, the cycle at which fluorescence minus the baseline reach the established threshold is identified as the threshold cycle (C_t). Nine data points were added between each experimental point using cubic interpolation in order to add decimals to the C_q. It is required to add data points by interpolation between experimental points in order to increase precision since gene copies increase roughly by two-fold between two experimental data points. This operation was performed using the *resize* MATLAB function. Additional file 4 present the threshold method MATLAB script.

Description of the SGSD method

Before applying the SGSD algorithm, data points were added by cubic interpolation between experimental data points. The number of points to be added was optimized using factorial Designs of Experiment (DoE), details of which are presented below. After adding data points, the amplification curves were smoothed and differentiated.

The Savitzky-Golay smoothing and differentiation method (SGSD) is based on mathematical treatment of spectral bands using convolution coefficients [Savitzky et Golay, 1964; Schafer, 2011]. In summary, the SGSD method uses a moving window of points to re-align each data point according to its neighbors, which reduces noise and allows further treatment such as differentiation. First, a polynomial of order N is fitted to the data in the window of width W centered on the data point being calculated. The center point is then repositioned according to the fitted polynomial, the polynomial is differentiated and the differentiated value of the center point is retained to construct the differentiated curve. The window is then moved over by 1 point to calculate the next point with the same approach, until the whole curve has been re-fitted and differentiated. The new differentiated curve is truncated at the beginning and at the end by $(W-1) \times 0.5$ points due to the impossibility to fit the polynomial for those points. Figure 2 shows how SGSD affects the differentiated PCR curve.

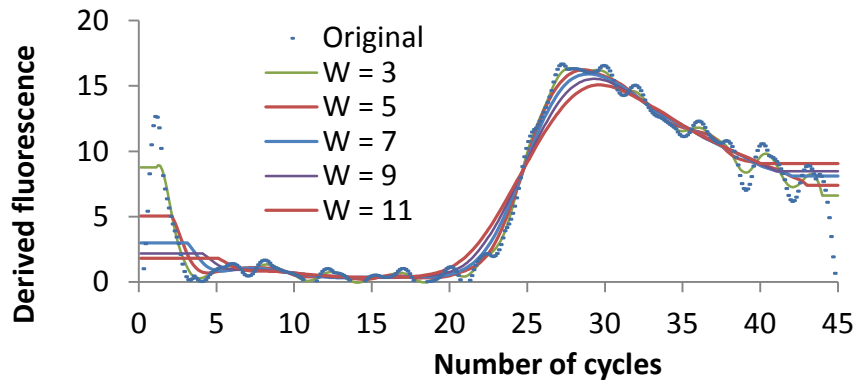


Figure 2. Effect of Savitzky-Golay smoothing on the 1st order differentiated PCR curve using width from 3 to 11 and a first order polynomial fit. The input DNA is 1000 gene copies.

This operation is realized by calculating convolution coefficients specific to the smoothing polynomial order, derivative order and the number of data point used to smooth each point. Briefly, these coefficients are obtained in two steps. First a Jacobian matrix J of $2M+1$ row with $N+1$ column with values m^n where M is the half-width, N the polynomial order, were $-M \leq m \leq M$ and $0 \leq n \leq N$ is created [Schafer, 2011]. Then convolution coefficient matrix (C) is obtained by solving the equation 9.

$$C = (J^T J)^{-1} * J^T \quad (9)$$

Each row of C have $2M+1$ number of coefficient used to smooth the spectral band and each different row correspond to different derivative power. Each individual smoothed data points are obtained by summing up the weighted adjacent data points inside the smoothing window. Additional file 5 present an example of calculations required for smoothing and additional file 6 describes the SGSD method MATLAB script.

After applying SGSD to the fluorescence curves, the cycle of quantification was found at the derivative maximum. The parameters that need to be adjusted for the SGSD method are the polynomial and derivative order, and the window width for the smoothing. These parameters were optimized using factorial designs of experiment (DoE) and details are presented subsequently.

In order to optimize the SGSD method, two factorials designs of experiments (DOE) were made to evaluate the effects of the method's parameters on its performance. The evaluated parameters were: number of points added by cubic interpolation, derivative order, polynomial order and width of the smoothing window. The analyzed responses were the bias, error and RSD score. The score for error was defined as the inverse of the weighted sum of global assay, inter-assay and intra-assay error, see equation (10). High scores denote high performance. Scores results were analysed using STATISTICA software in order to find significant optimums.

$$Score\ i = \frac{1}{\frac{Global\ i}{\sum_N^1 Global} + \frac{Intra\ i}{\sum_N^1 Intra} + \frac{Inter\ i}{\sum_N^1 Inter}} \quad (10)$$

Method comparison criteria

The algorithms were compared based on three primary criteria: relative standard deviation (RSD), relative bias (RB) and relative error (RE). These criteria were calculated using the gene copy estimation rather than the cycle of quantification as recommended [Bustin *et al.*, 2009]. RSD, equation (11, evaluates the methods' precision. RB, equation (12, detects systematic error. RE, equation (13, reflects methods' accuracy.

$$RSD = \sqrt{\frac{1}{N} \sum \left(\frac{y-\mu}{\mu} \right)^2} \quad (11)$$

$$RB = \frac{1}{N} \sum \left(\frac{x-y}{y} \right) \quad (12)$$

$$RE = \frac{1}{N} \sum \left\| \frac{x-y}{y} \right\| \quad (13)$$

Moreover, the methods were compared with two secondary criteria : residual normality and effect of competimers. All results were compared using Boneferroni-Holm correction for group comparison.

Normality of the residuals from the amplification reactions results was used to demonstrate their statistical predictability and to detect outliers. For a quality control tests, statistics can identify the false positive and false negative rate for a known population with defined specification limits. The error between the statistical predicted rate and the actual rate will depend on the residual normality. For each compared method, a straight line was constructed using residuals within ± 1 z-score. Distance from the straight line was calculated for each residual within ± 3 z-score and the SD (σ) was computed. Residuals at a distance of 1.96σ and more from the straight line were considered outlier. Residuals were also evaluated for their ability to match a 95 % limit prediction. Residuals located at z-score = 1.96 were compared to statistical prediction of the lower and upper 95 % residual limit (mean $\pm 1.96\sigma$).

As an example of quantification cycles and genes copies calculation by MATLAB using any of the four methods cited above, a script is available (additional file 7).

Results and discussion

SGSD parameters optimization

Two designs of experiment (DoE) were realized to investigate the following parameters effects on RE, RB and RSD: number of interpolation points, window width, polynomial order and derivative order. As a consequence of constraints inherent to the convolution coefficients calculations (table 1), more than one couple of polynomial and derivative order lead to the same differentiation results. Therefore, two DoE were used to complete the investigation. Results were analysed using the STATISTICA software, by ANOVA based on quadratic models with double interactions. Models were reduced to exclude non significant interactions.

Table 1. Savitzky-Golay constraints

Constraints		Variables definition	
1	$2M+1 \geq 3$	W	Width (2M+1)
2	$N \leq 5$	N	Polynomial order
3	$N \leq 2M$	M	Half-width
4	$d \leq N$	d	Derivative order

In the first DoE, the objective was to evaluate the effect of data point addition by interpolation, the derivative order and the smoothing width in order to minimize RE, RB and RSD. A 3^3 full factorial design was prepared as presented in table 2.

Table 2. DoE 1 – Level used for each parameter

	Level 1	Level 2	Level 3
Interpolation	None	3 pts	11 pts
Derivative and polynomial order	1st	2nd	3rd
Width	7 pts	9 pts	11 pts

The adjusted R^2 of the ANOVA models for RSD, RB and RE score responses are respectively 0.96, 0.92 and 0.89. P-values from the three ANOVA tables are reported in a summary, table 3. The summary shows that derivative order, width, and their interaction are significant, with the most important factor being the derivative and polynomial order. DoE results and complete ANOVA tables can be found in additional file 8.

Table 3. P-values obtained for DoE 1 ANOVA

Summary	RSD	RB	RE
A-Interpolation	1.13E-02	1.04E-02	1.12E-02
B-Derivative orders and polynomial order	1.23E-11	4.49E-10	5.10E-10
C-Width	1.78E-07	0.13E-05	6.17E-06
Interaction B-C	3.23E-04	3.69E-02	1.02E-02

Using multi linear regression coefficients to predict score values, desirability profiles (Additional file 8) and desirability surface (figure 3), were generated for RB, RE and RSD. The profiles suggest that derivative order should be set to 1; optimum smoothing width is larger than 11 points, therefore out of the experimental space; and a certain level of interpolation could be beneficial, although there is no clear statistical effect for this factor. Mean plot for interpolation, derivative order and smoothing width are shown in figure 4.

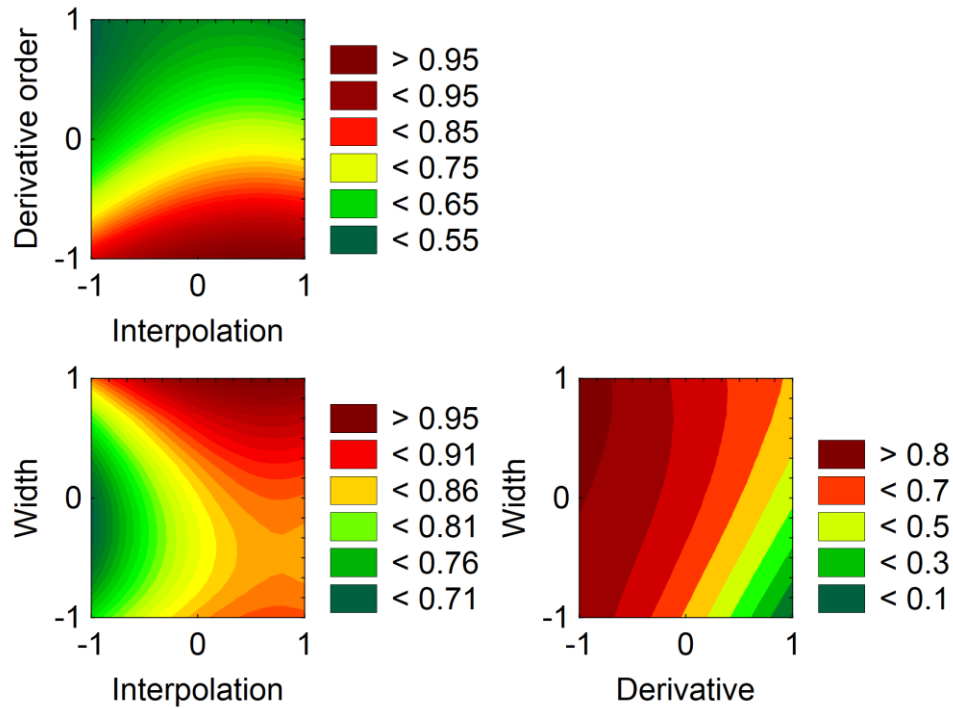


Figure 3. Desirability surface plot for interpolation, derivative order and smoothing width based on relative standard deviation, bias and error. The higher the desirability, the better (red=high, green=low).

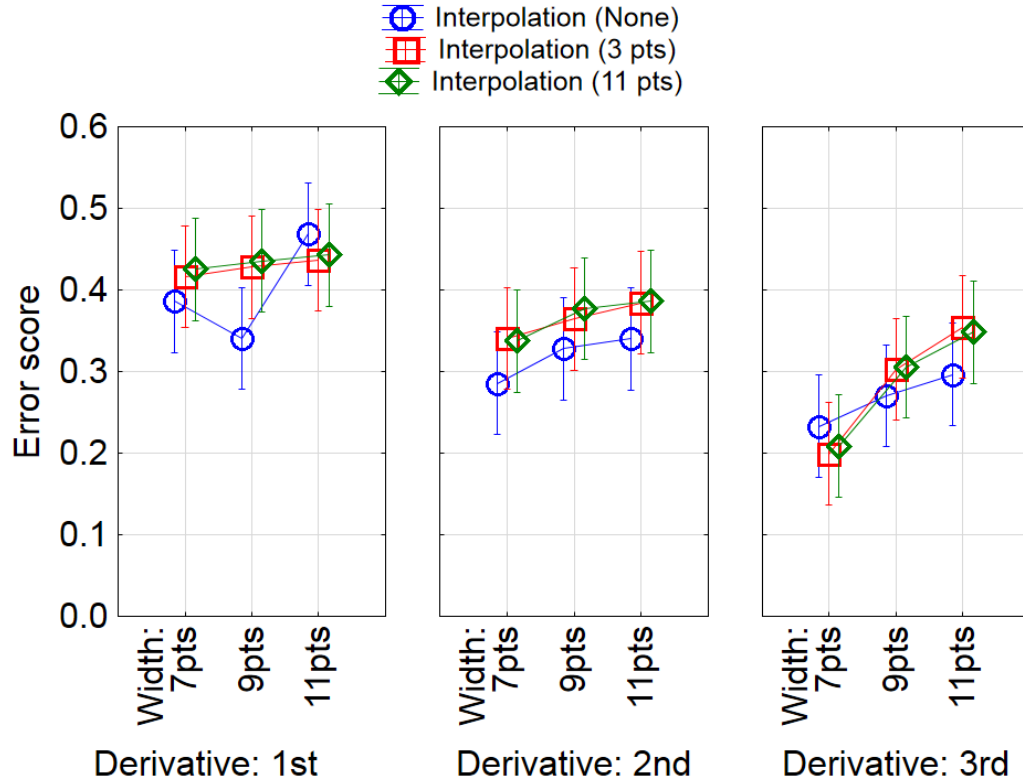


Figure 4. Mean plot for interpolation, derivative order and smoothing width.

The second DoE was prepared to investigate the effect of the polynomial order and evaluate further the interpolation and smoothing width. A 3^3 full factorial design was prepared with the levels detailed in table 4. The derivative order was set at first order during the experiment.

Table 4. DoE 2 – Level used for each parameter

	Level 1	Level 2	Level 3
Interpolation	7 pts	9 pts	11 pts
Derivative order	1 st	1 st	1 st
Polynomial order	1 st	2 nd	3 rd
Width	11 pts	13 pts	15 pts

The adjusted R^2 of those ANOVA models for bias, error and RSD response are all > 0.99 , indicating good fit of the model to the data. As shown in table 5, polynomial order and smoothing width are highly significant while data point addition by interpolation is only significant for RSD reduction. DoE results and complete ANOVA tables can be found in additional file 8.

Table 5. P-values obtained for DoE 2 ANOVA

Summary	RSD	Bias	Error
A-Interpolation	5.15E-03	3.22E-02	6.14E-02
B-Polynomial order	2.02E-12	2.88E-14	2.05E-13
C-Width	2.32E-03	4.20E-13	2.99E-13
Interaction A-B	1.31E-03	>0.05	3.41E-02
Interaction A-C	6.80E-07	3.17E-02	>0.05
Interaction B-C	1.01E-11	1.85E-15	1.13E-14

Observation of the desirability profiles (Additional file 8) and the mean plot of figure 5 suggest that polynomial order should be set to 1 for optimal results. Based on the two DoE conclusions, interpolation should be set at 7 points, although the effects are marginal. It is possible that the interpolation only affect the C_q 's decimal addition.

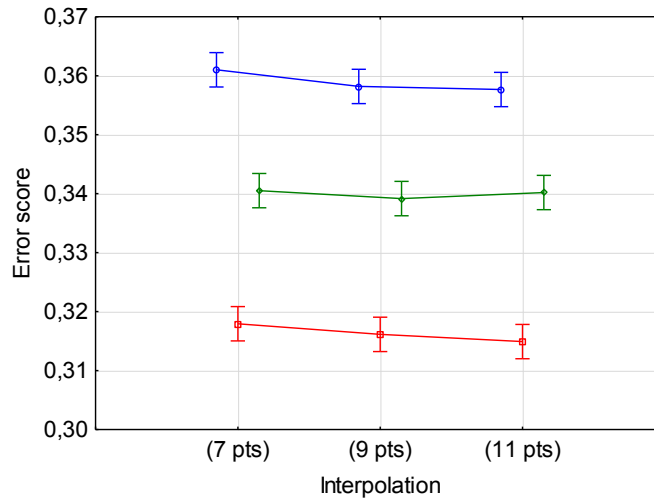


Figure 5. Mean plot for interpolation and polynomial order. First, second and third order polynomials are shown in blue, green and red respectively.

In order to obtain a firm knowledge of the optimal smoothing width, the tested range was increased in a separate test using widths varying between 2 and 19 experimental data points. Derivative order, polynomial order and data points addition by interpolation were set to their optimal values: 1, 1 and 7 respectively. The width (W) used for this analysis is based on experimental data points, not on the total number of points which include interpolated points. Thus it is understood that the window width including interpolated points (I) is equal to $1+(I+1)(W-1)$. Figure 6 illustrates the results, showing three local maxima. The third local maxima, width of 14.75, is the global maximum RE and RB. However, the calculated score in

the 3rd local maxima are not reliable (data not shown). The algorithm removed data points equal to $(W-1)/2$ at both end. At a width of 14.75, the loss of data affected the lowest DNA concentration results: if the C_q value is located in the edge where data is lost, the output C_q value will be pushed away from that frontier. At width of 14.75, the C_q would not be inferior to a certain value close to the smallest tested concentration which unfairly reduce the error, bias and standard deviation. In order to have a fair comparison between the different algorithms, the second local maxima, width of 9.75, was selected as the overall maximum.

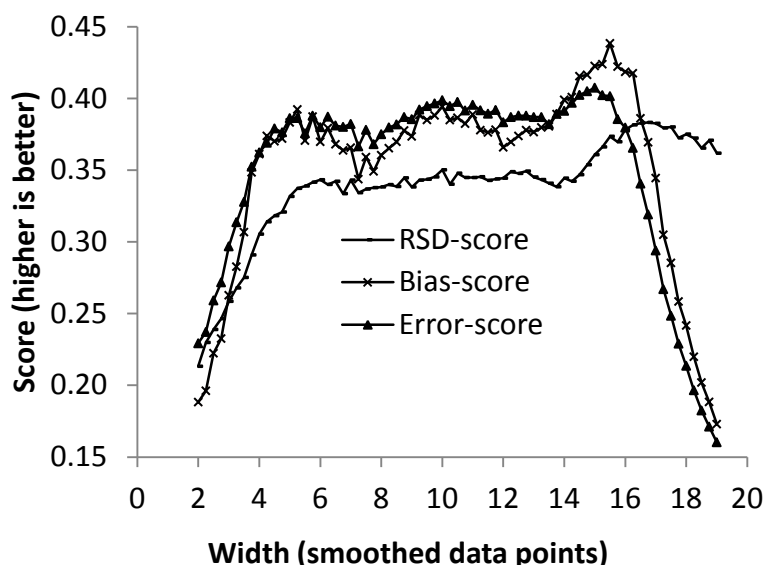


Figure 6. Effect of Savitzky-Golay smoothing width parameters on RSD, bias and error.

Smoothing width was found to be specific to the assay (data not shown). However, general guideline can be made toward width selection. A width ranging between 5 and 11 data points should give acceptable results. Below 5 data points, the smoothed curve is prone to inconsistent derivative maximum. Between 5 and 11 data points, truncations are made in the baseline and the plateau which should not affect significantly the results. The optimum should be determined by calculating error of prediction for standard DNA over a generous range of width.

Standard curves

Slopes, intercepts and R^2 are shown in table 6. Determination coefficient demonstrates similar linearity amongst investigated methods ($0.98 < R^2 < 0.99$). From the slope, the PCR reaction efficiency can be obtained by the following equation:

$$E = \left(10^{\frac{-1}{\text{slope}}} \right) - 1 \quad 14$$

Table 6. Standards curves parameters (n = 10)

	SGSD	4PLM	C_{y0}	C_t
Slope	-3.5 ± 0.1	-3.6 ± 0.2	-3.6 ± 0.2	-3,5 ± 0.1
Intercept	40.1 ± 0.5	37.9 ± 0.5	36.5 ± 0.5	35.6 ± 0.5
R²	0.98 ± 0.01	0.98 ± 0.01	0.98 ± 0.01	0.98 ± 0.02

The efficiencies (E) are reported in figure 7 and were compared together. Results show that the four methods belong to two efficiency groups. SGSD and C_t would belong to the first group ($E = 0.93$) while 4PLM and C_{y0} would belong to the second group ($E = 0.90$). Within groups efficiency variations are not statistically different (p-value = 0.36 and 0.18 respectively). However the two groups are different (p-value = 0.008). Amplification curve modeling could explain the efficiency difference. In the first group, quantification is performed using real or smoothed data points, while the second group quantification is performed using sigmoid curve modeling. Overall, the analysis of standard curves shows that the SGSD method performs similarly to other methods in term of efficiency determination using standard curve as a mean of absolute quantification.

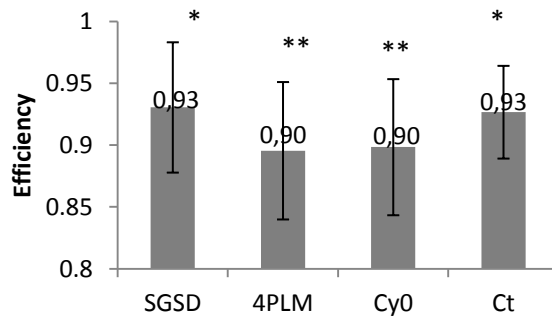


Figure 7. PCR mean efficiency (E) of standards curves

Comparison of quantification methods: global assay

Results for relative standard deviation (RSD), bias (RB) and error (RE) for the 200 PCR curves are found in figure 8. These results express the performance of the four algorithms against multiple uncontrolled factors coming from intra- and inter-assay variations. The RSD of SGSD is the lowest amongst compared method but is only statistically different to the C_t method (p-value = 0.004). Bias and error are significantly lower for the SGSD method (p-values respectively <0.005 and <0.01). The global performance of SGSD therefore appears to be best of the four investigated methods.

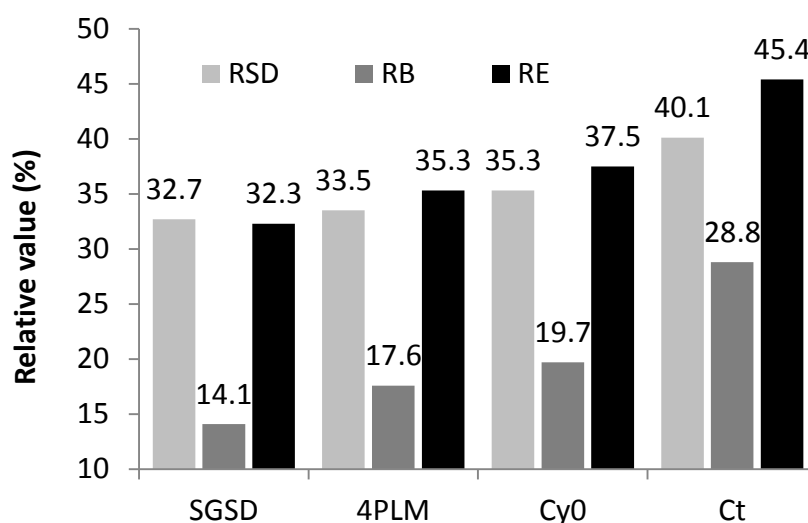


Figure 8. Global algorithm comparison for standard deviation, bias and error on a relative scale.

Comparison of quantifications methods: intra-assay

PCR reactions are influenced by various factors regrouped in two categories: intra- and inter-assay. Intra-assay variations result from mostly physical factors that vary between replicates of a PCR batch run, such as temperature, well cleanliness, physical obstruction and alteration of light path for fluorescence measurements. For perfect replicates, chemistry should be the same, but experimentally, slight variations are expected. Intra-assay variations reveal how the algorithm is able to perform with those varying factors. They are calculated by first dividing each G_c prediction by its input G_c . Then, all of the ten samples are averaged over the 20 experiments and means are compared to each other.

Figure 9 shows RB, RE and RSD for the four methods. SGSD, 4PLM and C_y0 have similar RSD and RE (p-value > 0.05). SGSD has different RSD and RE value than the C_t method (p-value = 0.02 and 0.005 respectively). Therefore, it can be concluded that SGSD, 4PLM and C_y0 performances are similar in a homogenous PCR experiment.

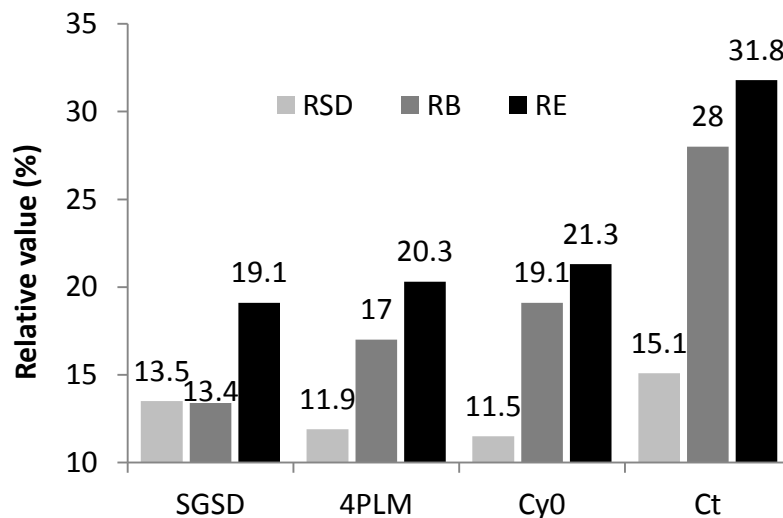


Figure 9. Intra assay comparison for bias (RB), error (RE) and standard deviation (RSD) on a relative scale.

Comparison of quantification methods: inter-assay

Inter-assay variations result from experimental variations and external factors such as reagent variations between batches, PCR inhibitors or activators, PCR thermocycler, different operators and variation in manipulation. Figure 10 give the RB, RE and RSD for the four methods. They are calculated by first dividing each G_c prediction by its input G_c . Then, mean of experiments or runs are compared to each other.

The SGSD method statistically outperforms the other methods in terms of error (p-value = 0.03, 0.05 and 0.004 respectively). It is not the case for RSD, for which the SGSD method is not statistically different than 4PLM and C_y0 (p-value = 0.11 and 0.12 respectively) but still outperforms C_t (p-value = 0.05). To demonstrate that RSD for SGSD is statistically better, a larger data set would be required since PCR variations are large and existing sampling is small (n = 20 experiments). Still, data suggest that the SGSD method performs better than the three other methods for multiple heterogeneous experiments.

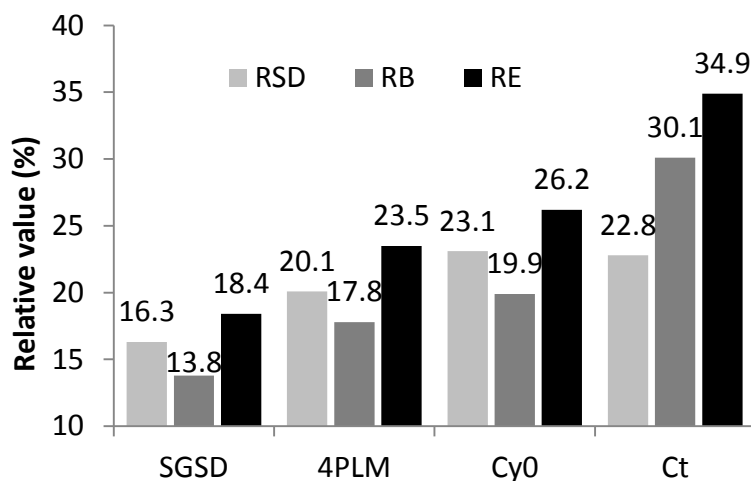


Figure 10. Inter-assay comparison: relative bias (RB), error (RE) and standard deviation (RSD).

Comparison of quantifications methods: summary

In summary, SGSD outperformed other methods on the global comparison because it was superior in inter-assay performance. The SGSD method is expected to perform similarly to other quantification method in homogenous experiments and better in experiments affected by multiple uncontrolled factors that vary from run to run.

Comparison of quantification methods: data normality

Using a normal probability plot (additional file 9) outliers were identified statistically and counted for each method (table 7). Results indicate that the SGSD and the C_y0 method have the fewest statistical outliers.

Table 7. Data points that deviate from the normality

	SGSD	4PLM	C _y 0	C _t
Line SD (-1 to 1)	0.10	0.12	0.08	0.15
Number of outliers (95% confidence)	5	12	6	8

Deviation from normality can be represented as in figure 11, where statistical prediction and actual observation of the residual 95 % confidence limit are compared. Positive limits (z-score = 1,96) are identified with the (+) sign and negative limits (z-score = -1,96) are identified with the (-) sign. According to these results, the SGSD and the C_y0 methods are the most statistically predictable, with lower and upper statistical limits differing slightly.

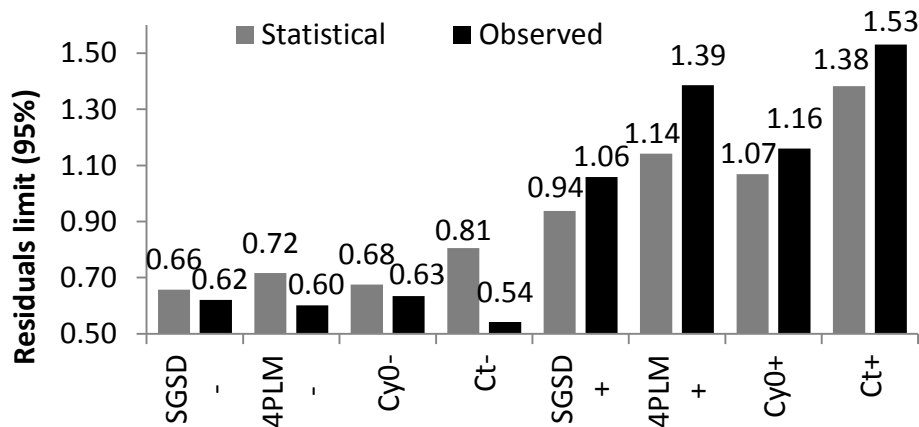


Figure 11. Observed residuals limit and the statistical residuals limit

Overall, the analyses of residuals and outliers demonstrate that the SGSD and C_y0 methods would be less affected by statistical outliers than the other tested methods. It would also be easier to establish statistical specification limits using the SGSD or C_y0 methods since their residuals are more normal than the other methods'. False positive rate would thus be lower for these two methods.

Comparison of quantification methods: Effect of competitors

When testing with increasing amounts of competitor, the standard deviation of the calculated C_q increases. The four method's precision (SD) was affected in a similar pattern, as shown by figure 12. Additionally, the C_q number increased proportionally with competitors and the SGSD method was distinctly less affected by the increase, as reported in figure 13. The C_q increased by 1.95 %, 2.50 %, 2.52 % and 2.25 % for each 1 % of competitor for, respectively, SGSD, C_y0 , 4PLM and C_t methods.

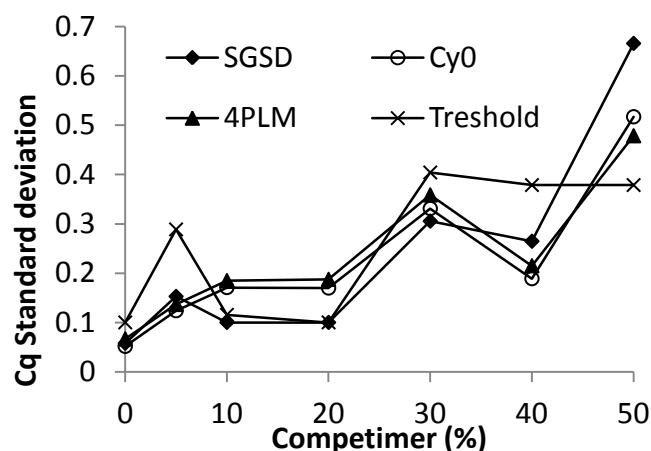


Figure 12. Effect of increasing concentration of competitor on Cq SD for the evaluated methods

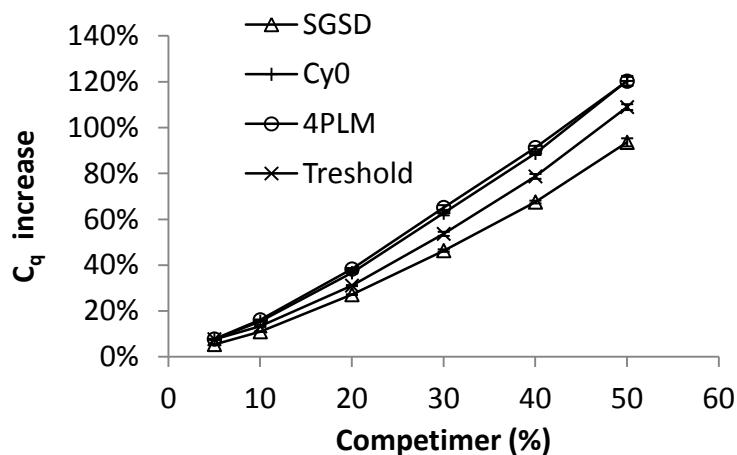


Figure 13. Effect of increasing concentration of competitor on the Cq for the tested methods.

Performance for different PCR assay

In order to validate the SGSD method on different PCR assay, data sets from the literature were used as described previously. Results for relative standard deviation (RSD), bias (RB) and error (RE) are shown in figure 14 (rounded and excluding C_t results). The C_t method showed an RSD of 30 to 33 %, RB of 4 to 17 % and RE of 21 to 24 %. All methods except the C_t method show statistically equivalent performances for all three criteria. This may results from the data sets coming from similar runs and which minimize the impact of inter assay variation. These variation were shown to be best controlled by the SGSD method. Thus the SGSD method was not challenged by these data. These results demonstrate that the SGSD method is valid and can be used with different assays.

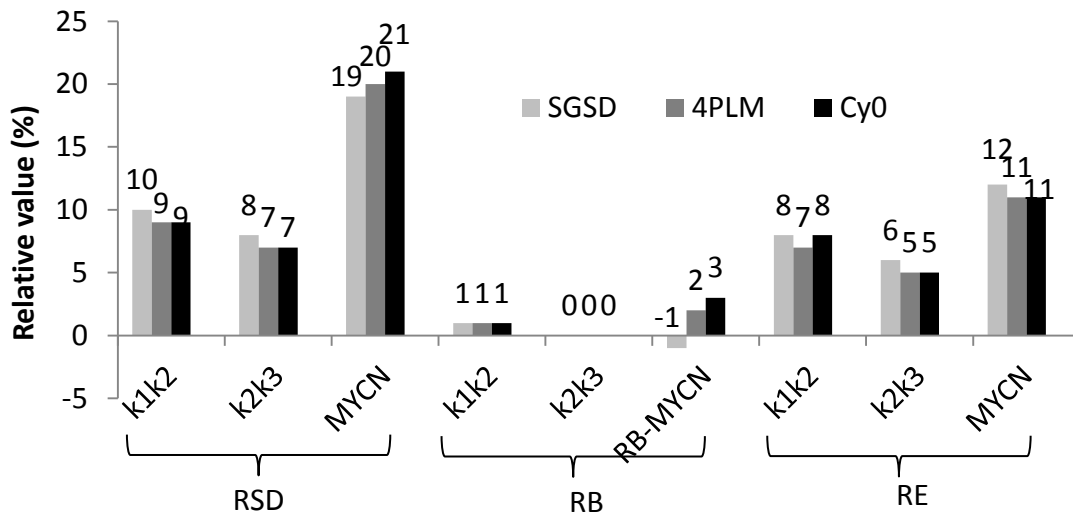


Figure 14. Evaluation of the SGSD, 4PLM and Cy0 method on three different assays. The target gene of the different assays are the k1k2 amplicon, the k2k3 amplicon and MYCN gene. The methods are evaluated based on relative standard deviation (RSD), bias (RB) and error (RE).

Conclusion

The Savitzky-Golay smoothing and differentiation (SGSD) method was adapted and optimized for PCR quantification. Optimal parameters found for the Savitzky-Golay smoothing are the addition, prior to quantification, of seven data points between each experimental points using cubic interpolation, first order derivative differentiation with first order polynomial fitting. The smoothing width parameter is specific to the assay and can be determined with standard curve data analyzes. However, a width ranging between 5 and 11 PCR points should yield good results.

The SGSD method was compared to two methods based on sigmoid curve fitting and to the traditional threshold method. The SGSD method did perform significantly better than other methods when challenged with inter-assay variations. The SGSD method performed similarly to sigmoid curve modeling methods on an intra-assay basis. Residual normality and residual predictability were superior when using the SGSD method. Competimer was found to affect the SGSD method's precision similarly as the other methods. However, the SGSD method seems to be less affected by the competimer C_q shift.

Results obtained from in-house data were validated using fluorescence data from the literature for intra-assay variation. On the other hand, inter-assay variation results could not be validated with other PCR chemistry since fluorescence data challenged for inter-assay variation could not be found in the literature. Overall results are sufficient to recommend its use, especially for assays submitted to inter-assay variations and noise. The SGSD method could also be combined with existing methods in order to evaluate whether existing method could benefit or not from smoothing of fluorescence data.

Authors' contributions

C. Gaudreault planned, executed and analyzed the qPCR experiments. C. Gaudreault also programmed MATLAB scripts for quantification from fluorescence data. C. Gaudreault analyzed results and wrote the manuscript. J. Salvas and J. Sirois supervised the work, the planning, contributed to idea generation and revised critically the manuscript.

Acknowledgements

We thank Ryan Gosselin from Université de Sherbrooke for MATLAB and Savitzky-Golay smoothing support and for useful suggestions. We thank Pfizer Montreal's microbiology lab team for giving us access to lab and equipments and for their wise advices. Part of this work was supported by the BMP-Innovation scholarship (NSERC/FRQNT) and by Pfizer.

Competing interests statement

The authors declare no competing interests.

Additional material

Additional file 1

MATLAB script for baseline adjustments, can be opened with Microsoft Word or with Notepad.

Additional file 2

MATLAB script for the four parameter logistic model (4PLM) quantification method, can be opened with Microsoft Word or with Notepad

Additional file 3

MATLAB script for the C_{y0} quantification method, can be opened with Microsoft Word or with Notepad

Additional file 4

MATLAB script for the threshold (C_t) quantification method, can be opened with Microsoft Word or with Notepad

Additional file 5

Microsoft Word file containing additional explanations about Savitzky-Golay convolution coefficient determination

Additional file 6

MATLAB script for the Savitzky-Golay smoothing and differentiation (SGSD) quantification method, can be opened with Microsoft Word or with Notepad

Additional file 7

MATLAB script example for C_q and G_q generation using any of the four methods cited above, can be opened with Microsoft Word or with Notepad

Additional file 8

Microsoft Word file containing a supplementary figures

Additional file 9

Excel file containing DoE and ANOVA results

7.4. Principales conclusions de l'article

Les conclusions de l'article sont que la méthode LDSG comparativement aux autres méthodes :

- Démontre une performance significativement supérieure, lorsque sujette à des variations inter-essais.
 - Erreur relative = 16.3%
 - Biais relatif = 13.8%
 - Écart-type relatif = 18.4%
- Démontre une performance similaire, lorsque sujette à des variations intra-essais.
 - Erreur relative = 13.5%
 - Biais relatif = 13.4%
 - Écart-type relatif = 19.1%
- Permet d'obtenir une meilleure normalité des résidus.
 - 5 données aberrantes sur 200
- Permet d'obtenir de meilleure prédiction statistique des résidus.
- Est moins affectée par les *competimers* au niveau du déplacement du C_q .
- Est affectée par les *competimers* au niveau de la précision au même niveau que les autres méthodes

L'article ne fait pas mention de l'algorithme du fournisseur pour des raisons de confidentialité. Cependant, les conclusions de cette section s'appliquent aussi à l'algorithme du fournisseur dans le sens où l'algorithme SGSD est plus performant puisqu'il a aussi été comparé durant l'étude des données, sans être inclus dans le manuscrit. Ainsi dans la section suivante, il sera vérifié si l'utilisation de l'algorithme LDSG en substitution de l'algorithme du fournisseur peut améliorer la précision (ETR) de la méthode. L'algorithme LDSG utilisera la première dérivée et un polynôme de degré 1 tel que décrit dans l'article. L'analyse des données suggérerait l'utilisation d'une fenêtre de lissage de 9,75 cycles de PCR pour une quantification allant jusqu'à 10 UCG. Cependant, considérant que la limite de quantification de la méthode est de 50 UCG, la fenêtre de lissage utilisé sera de 14,75 cycles de PCR. Ce choix de fenêtre fait en sorte que la quantification est erronée pour les quantités avoisinant les 10 UCG, mais augmente les performances pour les quantités d'UGC supérieur.

7.5. Impact sur la précision de la méthode de qPCR

7.5.1. Résultats

L'algorithme LDSG a été substitué à l'algorithme du fournisseur pour les essais de la version 1, 2 et 3 du protocole de la méthode tel que décrit dans la section 7.2.2. Les nouvelles valeurs de précision pour chaque version du protocole sont présentées dans le Tableau 7.1. La précision a été caractérisée en s'appuyant sur deux composantes : la composante globale et la moyenne des duplicatas de PCR. Les duplicatas de PCR sont des répétitions analytiques visant à augmenter la précision de la méthode. Étant donné que chaque échantillon d'ADN purifié est distribué dans deux puits de PCR différents et que le résultat de ces puits est moyenné dans la méthode, la précision de la méthodologie est donnée par l'analyse de la moyenne des duplicatas de PCR alors que la précision globale ne tient pas compte de cette opération mathématique. La distinction entre les deux types de précisions permet d'analyser l'effet direct du changement d'algorithme sur la précision sans être biaisé par l'atténuation de la variance due aux répétitions analytiques.

L'analyse des résultats permet de voir que l'utilisation du nouvel algorithme permet de réduire écart-type relatif (ETR) global de 9 % pour la première version du protocole. Cette différence n'est pas statistiquement significative, probablement en raison du faible échantillonnage. Cependant, l'analyse des duplicatas de PCR moyenné donne une différence de 4 % statistiquement significative ce qui tend à démontrer que la différence globale est réelle. L'analyse des données du protocole #2 démontre une réduction globale statistiquement significative de 4 et 3 % pour l'ETR globale et de la méthode respectivement. L'analyse des données du protocole #3 ne montre aucune différence au niveau de l'ETR globale et une augmentation de 1 % statistiquement non significative pour le nouvel algorithme. Ainsi, pour le dernier protocole en vigueur, le nouvel algorithme n'aurait aucun effet.

Tableau 7.1. Effet des différents algorithmes sur les tests au niveau de la précision

Version du protocole	Analyse de la précision -Écart type relatif (ETR)	Algorithme Fournisseur (ETR %)	Algorithme LDSG (ETR %)	t-test** (valeur p)
Version 1, 2012 (n=17)	Global	68,2	59,4	0,0794
	Méthode *	57,4	53,1	0,0522
Version 2, 2012 (n=104)	Global	47,5	43,2	0,0030
	Méthode *	42,5	39,1	0,0332
Version 3, 2013 (n=52)	Global	38,1	38,4	0,4080
	Méthode *	32,2	33,4	0,1340

* La méthode utilise les moyennes des répétitions analytiques (duplicatas)

** Analyse utilisant une distribution unilatérale; tests effectués par paires

7.5.2. Analyse

L'effet de l'algorithme sur la méthode au niveau de l'erreur et du biais n'a pas pu être évalué. Pour être en mesure de calculer l'erreur et le biais, tel qu'effectué dans l'article, il aurait fallu connaître les valeurs réelles d'UCG par échantillons ce qui est difficilement réalisable. L'utilisation des UFC comme valeur de référence n'aurait pas été possible à cause des différences intrinsèques des UFC et des UCG décrites aux sections 3.2 et 4.1.3. De plus, le but initial ayant amené ces travaux à s'intéresser à l'algorithme est d'augmenter la précision.

La différence de performance la plus marquée entre les deux algorithmes est au niveau de la première version du protocole. Cette version avait le défaut de co-purifier des substances inhibitrices de PCR avec l'ADN. Ainsi, les résultats suggèrent que l'algorithme LDSG est plus robuste que celui du fournisseur lorsque soumis à des substances ayant le potentiel de causer des problèmes d'inhibition de PCR. Les problèmes mineurs d'inhibition de PCR ont pour effet d'ajouter du bruit sur la courbe de PCR, il est possible que l'effet du lissage de l'algorithme LDSG atténue le bruit causé par les problèmes mineurs d'inhibition, ce qui pourrait expliquer l'avantage de celui-ci pour l'analyse de données sujettes à l'inhibition.

La problématique d'inhibition étant réglée, la précision de l'algorithme LDSG est toujours statistiquement supérieure à celui du fournisseur pour la seconde version du protocole. On s'attendait à ce que ces gains de précision se maintiennent pour le troisième protocole, ce qui n'est pas le cas puisque leur performance est similaire. Les changements de protocole entre la

version 2 et 3 du protocole sont strictement procéduraux et ne devraient pas affecter les performances de l'algorithme. Ces changements concernent le nettoyage d'une pièce d'équipement réutilisable, le temps de centrifugation, l'homogénéisation de l'ADN et du master mix dans les puits de PCR et la préparation des bactéries. Cependant, les lots de réactifs ont changé ce qui a potentiellement changé légèrement la chimie des réactifs. Il est probable que le fournisseur ait augmenté la robustesse de ses réactifs. Comme l'algorithme LDSG performe de façon similaire aux autres types d'algorithme lorsque les variations inter-essais sont minimales et que ses performances sont supérieures aux autres méthodes quand les variations inter-essais sont grandes, on suppose que les résultats observés entre le protocole 2 et le protocole 3 sont expliqués par l'augmentation de la robustesse découlant des réactifs. L'utilisation de réactifs plus robustes aurait pour effet de diminuer les différences inter-essais.

7.6. Sommaire

Dans le but d'augmenter la précision de la quantification, un algorithme de quantification basé sur le lissage et la différentiation de Savitzky-Golay (LDSG) a été développé. L'algorithme a été optimisé et comparé à d'autres algorithmes de quantification provenant de la littérature. L'algorithme LDSG a démontré des performances supérieures, lorsque soumis à des variations inter-essais au niveau de la précision, l'erreur et le biais.

L'algorithme LDSG a été utilisé en remplacement de l'algorithme du fournisseur pour les tests de performance analytiques afin d'augmenter la précision, définie par l'écart type relatif entre les résultats, de la méthode. L'analyse de la précision a démontré que l'algorithme était plus robuste en condition d'inhibition de PCR. Pour des raisons indéterminées, l'algorithme s'est montré supérieur à l'algorithme du fournisseur pour une version du protocole et similaire pour la version finale. Une cause potentielle serait la variabilité des réactifs de PCR et l'avantage de LDSG d'être robuste aux variations de réactifs. Ainsi, le nouvel algorithme n'augmentera pas la précision du protocole final, cependant, la robustesse de la méthode serait relevée. Une robustesse plus grande permet à la méthode d'accommoder une plus grande variabilité au niveau de la composition chimique des échantillons et des réactifs. En tolérant une plus grande variabilité, la validation de la méthode devrait être plus facile. Aussi, une robustesse accrue, peut accroître la gamme de produit pouvant être testés.

8. ADAPTABILITÉ DE LA MÉTHODE

8.1. Mise en contexte

Le protocole expérimental a été conçu et optimisé par le fournisseur pour l'analyse de l'eau purifiée. Pour évaluer le potentiel de la méthode pour l'analyse d'échantillons pharmaceutiques, on doit d'abord connaître pour quel type d'analyse la méthode peut être utilisée. Par type d'analyse on fait référence aux tests sur l'eau, les liquides, les poudres et les comprimés. L'objectif de ce projet n'est pas d'effectuer du développement de méthodes. Ainsi, dans cette section, on évaluera pour quel type de matière la méthode peut être rapidement adaptée sans modifier de façon drastique le protocole.

8.2. Méthodologie

8.2.1. Poudres et comprimés

Les produits solides sous forme de comprimés doivent être préalablement broyés, comme dans le cas de la méthode traditionnelle, afin d'obtenir une poudre avant d'être traités par le protocole. La méthode utilisée pour le broyage consiste à placer les comprimés dans un mortier stérile pour ensuite broyer ceux-ci à l'aide d'un pilon stérile. Dans le cas des matières premières, l'étape de broyage n'est pas nécessaire puisque ces matières sont déjà sous forme de poudre. La stratégie d'adaptation du protocole consiste à ajouter la poudre dans l'eau pour que les bactéries passent de la poudre à l'eau. Dans le cas d'une poudre totalement solubilisée dans l'eau, le total de la masse bactérienne, initialement dispersée dans la poudre, se retrouvera dans l'eau. Dans le cas d'une poudre insoluble, les cellules se déplaceront du solide vers le liquide à la manière d'une extraction solide-liquide grâce à l'agitation.

8.2.2. Adaptation pour matières non filtrable

Le principal obstacle à l'adaptation du protocole est le colmatage à l'étape de filtration pour les matières non filtrable. Ainsi, des ajouts au protocole auront pour but de résoudre cette problématique. Les ajouts ou modifications proposées sont la sédimentation, la pré-filtration, la dissolution acide et l'utilisation d'une petite quantité de matière pour les analyses. Les changements à la méthode pour l'analyse de matières sont fonction du type de matière et se présentent sous forme d'un arbre de décision présenté dans la Figure 8.1.

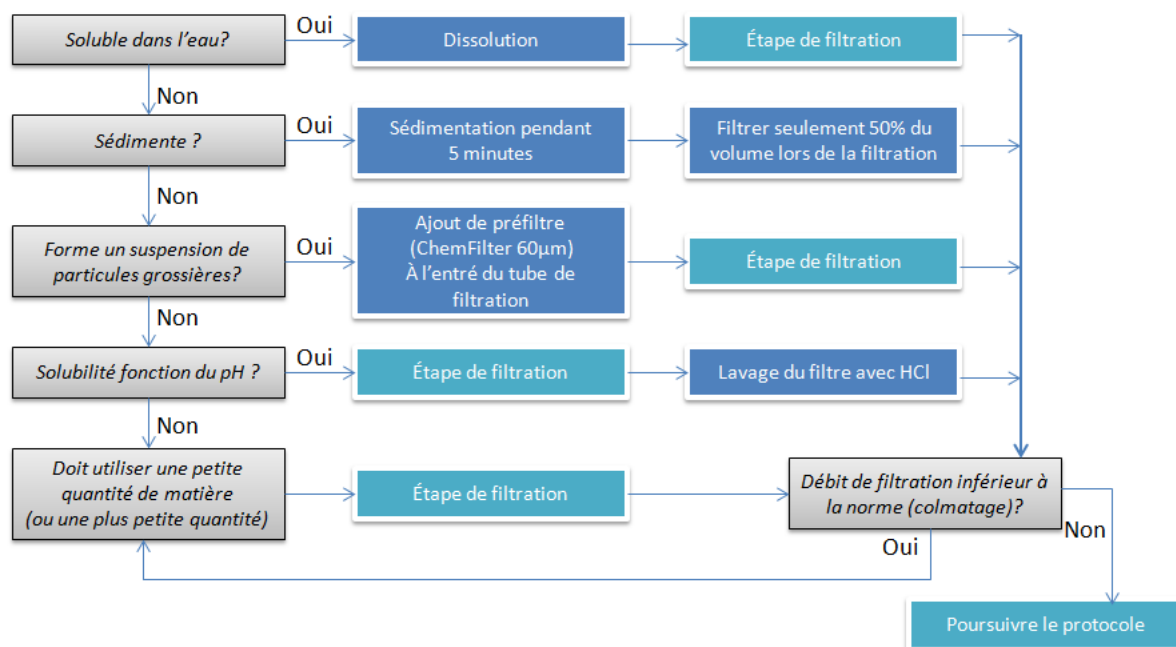


Figure 8.1. Arbre décisionnel pour le traitement des troubles de colmatage du filtre

Notez que l'utilisation d'une très petite quantité de matière peut permettre l'utilisation du protocole avec n'importe quelle matière. Cependant lorsque la quantité de matière testée est au-dessous d'une certaine quantité, les résultats rapportés ne sont plus d'aucune utilité puisque le bruit de fond est amplifié à un seuil qui est supérieur à la spécification de cette matière. L'équation 8.1 de la section 8.2.4, permet de déterminer la quantité de matière minimale pouvant être testé.

8.2.3. Élimination de l'ADN extracellulaire

Il est fortement probable que les matières pharmaceutiques contiennent de l'ADN extracellulaire. Lors de la mise en solution de la matière, l'ADN extracellulaire se retrouvera dans le solvant, l'eau. Environ 200 à 300 µl de solvant mouille le filtre et se retrouve dans les étapes de purification d'ADN. Ainsi, si cet ADN extracellulaire est en concentration élevée dans les 200 à 300 µl de solvant mouillant le filtre, les résultats seront faussés puisque cet ADN sera purifié et quantifié. Donc, pour éviter qu'une concentration élevée d'ADN extracellulaire fausse les résultats, une étape de rinçage du filtre avec de l'eau purifiée a été

ajoutée après l'étape de filtration de l'échantillon afin d'éliminer ce type d'ADN des étapes subséquentes.

8.2.4. Détermination de la quantité de matière utilisée

La réglementation USP <61> demande à ce que les tests soient effectués sur un échantillon de 10 g ou 10 ml. Dans le cas de la méthode à l'étude, 10 g représente une quantité pouvant engendrer des problèmes de colmatage pour la majorité des matières, sous forme de poudre, à tester. Ainsi, dans la plupart des cas, 10 grammes pourront être prélevés d'un échantillon représentatif du matériel à tester, puis dilués dans une solution aqueuse. La solution permettra un transfert des bactéries de la poudre vers la solution. Un échantillon de cette solution pourra ensuite être prélevé et on considèrera que la concentration bactérienne de l'échantillon est représentative de la solution complète dans laquelle les 10 grammes échantillonnés ont été dilués. Ainsi, pour un échantillon de 10 g dilué dans 200 ml de solvant pour lequel 10 ml est prélevé pour l'analyse, la méthode de qPCR analysera l'équivalent de 0,5 g de matière. Il est entendu que les 10 ml prélevés seront ajoutés dans un récipient et que le volume sera ajusté à 100 ml pour satisfaire la première étape du protocole étant la filtration de 100 ml de fluide. Les résultats seront reportés par gramme, ou millilitre, de matière.

La quantité de matière minimale qui doit être analysée par la méthode de qPCR est fonction des spécifications de la matière à tester et du bruit de fond de la méthode pour l'analyse de cette matière. Ainsi le bruit de fond rapporté par gramme de produit est fonction de la quantité de produit testé. Donc, plus la quantité testée est petite, plus le bruit de fond présent sera amplifié par l'effet multiplicatif de la dilution lorsque rapporté au volume complet de l'échantillon. Par exemple, si le seuil d'acceptabilité, aussi appelé spécification, pour un produit est de 500 UFC (ou UCG) par gramme, que le bruit de fond est de 50 UFC par essai et que 0,1 gramme de produit est testé, le bruit de fond corrigé sera alors de 500 UFC par gramme, soit l'équivalent du seuil d'acceptabilité. Les résultats n'ont alors aucune crédibilité. On peut ainsi déduire que la quantité minimale pouvant être testée pour un produit est donnée par la formule suivante:

$$m = \frac{LOQ_I}{S} = \frac{B + 3\sigma}{S} \quad 8.1$$

où m est la masse de matière minimale pouvant être testée, LOQ_I est la limite de quantification inférieure, S est la spécification de la matière, B le bruit de fond et σ l'écart type de celui-ci. Tel que discuté dans la section 5.3.5, la LQ_I proposée est de 3 écarts-types au-dessus du bruit de fond.

Prenez note que pour chaque matière différente et chaque variation au protocole, le bruit de fond et son écart-type devront être calculés.

8.2.5. Matières à l'essai

L'adaptation du protocole a été tentée sur plusieurs matières différentes. Le choix des matières est basé sur le nombre de tests annuels effectué sur ceux-ci, la représentativité des types de matière pouvant être testés et leurs disponibilités à la pharmacie de l'usine. Le protocole expérimental a été adapté pour les tests de matière, mais n'a pas été optimisé dans tous les cas. Le but est d'évaluer quel type de matière sera facile à analyser avec la méthode et quel type de matière représentera un défi. Les matières à l'essai sont répertoriées dans le Tableau 8.1.

Tableau 8.1. Matière à l'essai pour l'adaptation de la méthode

Matières solides		Matières liquides
Xylitol	Amidon NF/EP	Huile de poisson
Cellulose microcristalline	Amidon pré-gélatine	Sirop de maïs
CaCO ₃ (5 % maltodextrine)	CaCO ₃ (trace d'amidon)	Produit liquide #1 (sirop)
Ginseng	Vitamin D3	Produit liquide #2 (sirop)
Produit solide #1 (1 ingrédient actif)	Produit solide #3 (formulation minérale)	Produit liquide #3 (liquide pour vaporisateur)
Produit solide #2 (Vitamines et minéraux)	Produit solide #4 (1 ingrédient actif)	Produit liquide #4 (suspension)

8.3. Résultats

8.3.1. Matières solides

Sur les douze matières solides à l'essai, la méthode n'a pu être adaptée que pour cinq matières. Les adaptations requises pour chaque matière ainsi que les matières non adaptées sont résumées dans le Tableau 8.2.

Tableau 8.2. Adaptation du protocole pour les matières solides

Matières	Dissolution aqueuse	Sédimentation	Préfiltration	Dissolution du gâteau (HCl)	Tester une petite quantité	Matières non adaptées
Xylitol	X					
Cellulose microcristalline		X	X			
CaCO ₃ (5 % maltodextrine)			X	X		
Ginseng					X	
Produit solide #1					X	
Produit solide #2						X
Amidon pré-gélatine						X
Amidon						X
CaCO ₃ (trace d'amidon)						X
Vitamin D3						X
Produits solides #3						X
Produits solides #4						x

Xylitol

La méthode n'a pas besoin d'adaptation particulière pour l'analyse du xylitol. La solubilité du xylitol est de 163 g par 100 ml d'eau à 20°C [Wang *et al.*, 2007], ainsi la dissolution de la substance rend la solution de celle-ci apte à l'étape de filtration. L'analyse de cette matière est insensible au niveau du bruit de fond et aux spécifications puisqu'il est possible d'en tester une grande quantité.

Cellulose microcristalline

La méthode peut être adaptée pour l'analyse de la cellulose en utilisant une étape de sédimentation ou en utilisant une pré-filtration. L'analyse est possible en mettant 0,5 g de cette substance en suspension dans 200 ml d'eau et en laissant la poudre sédimenter. Ensuite, il suffit de pomper 100 ml de surnageant. Le résultat de quantification est ainsi rapporté par rapport à 0,25 g. Il est aussi possible de séparer la phase liquide du solide en utilisant une étape de pré-filtration avec un filtre de nylon de 60 μm . La séparation est soit due à la grosseur des particules ou à l'affinité chimique entre la cellulose et le filtre. Il est aussi possible d'utiliser une combinaison de sédimentation et de pré-filtration.

Carbonate de calcium avec trace de maltodextrine

La méthode demande une adaptation complexe pour l'analyse du calcium carbonate qui contient des traces de maltodextrine. Deux problématiques englobent l'analyse du CaCO_3 . Premièrement, le carbonate de calcium est insoluble à pH neutre, ne sédimente pas et colmate le filtre à bactérie. Deuxièmement, la maltodextrine est co-purifiée avec l'ADN et est probablement, comme plusieurs polysaccharides, un inhibiteur de PCR [Monteiro *et al.*, 1997; Schrader *et al.*, 2012]. L'effet combiné du carbonate de calcium et de la maltodextrine occasionne des problèmes d'inhibition de PCR et d'altération des lectures de fluorescence, Figure 8.2.

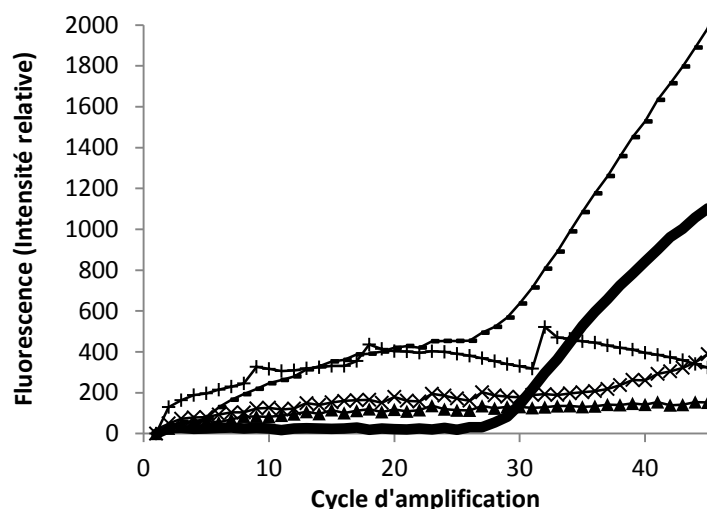


Figure 8.2. Effet combiné du CaCO_3 et de la maltodextrine sur les données de fluorescence du qPCR; amplifications affectées par la présence du produit (tiret, croix, axe, triangle) et amplification non exposée au produit (en gras)

Le problème de colmatage est réglé en lavant le filtre colmaté avec de l'HCl 0.05N, ce qui dissout le CaCO₃. Le lavage acide du filtre ne semble pas affecter les résultats des tests étant donné que la récupération des bactéries n'est pas affectée (Figure 8.3). La poudre de CaCO₃ est traitable jusqu'à 0,5 g, au-dessus de cette quantité, le lavage avec HCl devient une opération difficile. Considérant les difficultés de l'opération et les restrictions dues aux spécifications et aux bruits de fond, une quantité de 0,3 g serait idéale pour les essais. Le problème de la maltodextrine est réglé en pré-filtrant la suspension avec un filtre de nylon de 60 µm. On suppose que la maltodextrine adsorbe sur le filtre par affinité chimique.

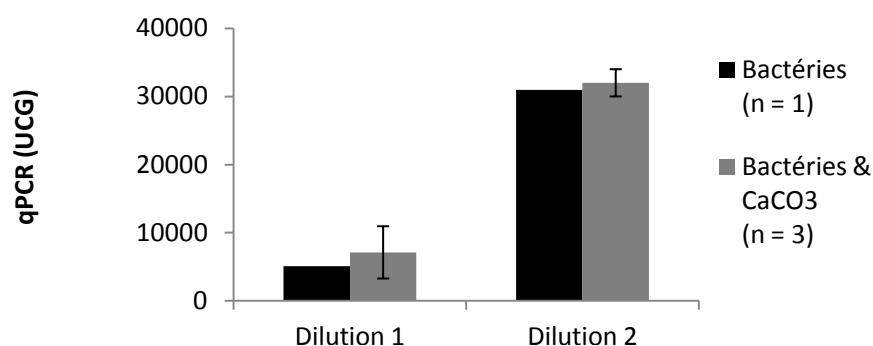


Figure 8.3. Récupération avec lavage acide

Ginseng et produits solides #1 et 2

Aucune modification simple du protocole ne permet l'analyse de ces matières. Les poudres sont insolubles et n'ont pu être séparées de la phase aqueuse. Ainsi, afin de tester ces poudres, de très petites quantités furent utilisées (0,05, 0,25 et 0,25 pour le ginseng, pour le produit solide #1 et pour le produit solide #2 respectivement) afin de prévenir le colmatage du filtre à bactérie.

Matières non adaptées

Certaines matières n'ont pu être adaptées pour être traitées par la méthode. Les deux types d'amidon testés étaient insolubles dans l'eau, colmataient le filtre à bactéries et ne pouvaient être séparés du solvant. Le produit solide #3 qui contient ces types de carbonate de calcium, ainsi que la vitamine D3 et le produit solide #4 ont présenté les mêmes problèmes. Afin de tester ces matières, il aurait été possible d'utiliser une très petite quantité de matières, cependant, le bruit de fond aurait été amplifié de façon à masquer le signal.

8.3.2. Matières liquides

Sur les six matières solides à l'essai, la méthode a pu être adaptée pour quatre matières. Les adaptations requises pour chaque matière ainsi que les matières non adaptées sont résumées dans le Tableau 8.2.

Tableau 8.3. Adaptation du protocole pour les matières liquide

Matières	Solution homogène	Sédimentation	Préfiltration	Dissolution du gâteau (HCl)	Tester une petite quantité	Matières non adaptées
Huile de poisson					X	
Sirop de maïs	X					
Produit liquide #1	X					
Produit liquide #2	X					
Produit liquide #3						X
Produit liquide #4						X

Sirop de maïs

Le sirop de maïs est une solution aqueuse qui pourrait à priori être filtrée directement. Cependant, la viscosité du sirop de maïs augmente la pression requise pour la filtration et rend celle-ci difficile. De plus, on ne sait pas si la viscosité de la solution est compatible avec les étapes subséquentes de la méthode. Ainsi le sirop de maïs est d'abord dilué avec de l'eau dans un rapport sirop:solvant de 1:3. L'analyse de cette matière est insensible au niveau du bruit de fond et aux spécifications puisqu'il est possible d'en tester une grande quantité.

Produit liquide #1 et 2

La méthode n'a pas besoin d'adaptation particulière pour l'analyse des produits liquides #1 et #2 puisque ceux-ci sont miscibles et filtrables. Ces matières auraient pu être ajoutées directement au filtre à bactéries au lieu d'être mélangées à l'eau. Cependant il est préférable de les diluer, car ces produits laissent des traces de colorant sur le filtre s'ils sont ajoutés directement à celui-ci, ce qui n'était pas le cas si un mélange était fait avec l'eau. L'analyse de cette matière est insensible au niveau du bruit de fond et aux spécifications puisqu'il est possible d'en tester une grande quantité.

Huile de poisson

Aucune modification simple du protocole ne permet l'analyse de ces matières. Puisque certaines composantes de l'huile de poisson s'accumulent sur le filtre, l'utilisation d'une petite quantité de cette matière permet l'utilisation du protocole.

Produit liquide #3 et #4

Ces deux produits finis liquides n'ont pu être adaptés au protocole puisque certaines composantes colmataient le filtre. Il n'était pas possible d'utiliser une très petite quantité de ces produits pour les analyses puisque la quantité requise aurait amplifié le bruit de fond au-delà du signal. L'analyse des matières premières constituant ces deux produits auraient permis de savoir quelle matière cause le colmatage.

8.4. Discussion

8.4.1. Matières non adaptées et pistes de solutions

L'adaptation de la méthode pour l'analyse des diverses matières discutées dans ce chapitre ne concerne que les modifications au protocole qui peuvent être rapidement mises en place et testées. Il est entendu que pour chaque matière, il existe probablement une méthodologie qui permettrait d'effectuer une extraction et purification d'ADN avec succès.

Centrifugation

Pour plusieurs substances, il est probable qu'une centrifugation aurait suffi à faire sédimenter les particules insolubles. La centrifugation n'a pas été testée dues aux risques d'y faire sédimenter les cellules avec les particules. Un travail d'optimisation pourrait permettre d'obtenir une sédimentation des particules les plus grossières, sans faire sédimenter les cellules. Cependant cette avenue n'était pas prioritaire dans l'ensemble des travaux.

Sédimentation

Au lieu d'utiliser de petites quantités de matière à tester, il aurait été possible de simplement faire sédimenter un produit et dilué son surnageant étant donné que certaines substances ne sédimenter que partiellement. Cette stratégie aurait permis de tester une plus grande quantité de matière que la stratégie des petites quantités.

Différences de propriétés et échange liquide-liquide

Dans certains cas, au lieu d'utiliser une petite quantité, comme pour l'huile de poisson, on aurait pu exploiter les différences de propriété. Dans le cas de l'huile, on aurait pu profiter de l'affinité de certaines substances pour la phase organique. Si on suppose que les substances qui colmataient le filtre avaient une meilleure affinité pour la phase organique, on aurait pu créer un mélange hétérogène huile-eau favorisant la retenue des matières problématique par la phase organique. Ainsi, on pourrait filtrer seulement la phase aqueuse et corriger le résultat de PCR par un facteur correctif considérant le partage des cellules entre la phase organique et la phase aqueuse.

8.4.2. Stratégie pour l'analyse de produits finis

Une approche systématique d'adaptation du protocole expérimentale pour un produit devrait inclure en premier lieu une adaptation pour les matières premières qui compose ce produit. Les produits sont généralement un mélange plus ou moins complexe de matière première. Ainsi, le développement de méthode pour l'analyse de produits devrait être grandement facilité si les méthodes d'analyse pour ses matières premières ont déjà été développées.

8.5. Sommaire

Au niveau de l'adaptation du protocole, on peut distinguer quatre catégories de matières classés en ordre de difficulté en terme de développement de méthode d'extraction d'ADN :

- **Catégorie #1**
 - Les liquides, compatible avec l'eau, contenant des composés solubles
 - Les poudres solubles
- **Catégorie #2**
 - Les poudres insolubles dont les particules de grande taille sédimentent facilement
- **Catégorie #3**
 - Les matières insolubles ne colmatant pas le filtre, lorsque dilué
- **Catégorie #4**
 - Les matières complexes qui comprennent plusieurs composés et/ou qui requièrent beaucoup d'effort afin d'adapter le protocole.

Sur les 18 matières testées, quatre appartiennent à la catégorie #1, une appartient à la catégorie 2, trois appartiennent à la catégorie #3 et dix appartiennent à la dernière catégorie. De la dernière catégorie, une seule matière, le CaCO_3 , a été le sujet de travaux pour être adapté au protocole. Ainsi, huit matières sur les dix-huit investiguées sont aisément traitées par la méthode de qPCR.

Le protocole est adapté pour les tests d'eau et de matières filtrables tels que les produits finis liquides #1 et #2. Il est aussi facilement adaptable pour les matières solubles ou compatibles avec l'eau comme le xylitol et sirop de maïs. Pour les matières insolubles de taille de particules élevées, telle que la cellulose, la décantation permet de générer un surnageant adéquat pour les analyses. Pour certaines matières telles que le CaCO_3 qui contiennent des traces de maltodextrine, l'adaptation du protocole est réalisable, mais complexe et ajoute de nombreuses manipulations. Pour certaines matières comprenant de nombreux composés, telle que le ginseng, la séparation des différentes composantes colmatant le filtre n'est pas pratique et il faut alors diluer le plus possible cette matière pour éviter le colmatage du filtre, sans amplifier le bruit de fond au-delà des spécifications pour cette matière.

9. INFLUENCE DES BACTÉRIES TOTALES

9.1. Mise en contexte

La méthode de qPCR à l'essai quantifie l'ADN des cellules bactériennes d'un échantillon filtrable, tel que décrit dans les chapitres précédents. L'ADN peut provenir de cellules vivantes capables de croître dans un milieu de culture, de cellules vivantes non cultivables (CVNC) ou encore de cellules mortes dont l'intégrité membranaire subsiste. Ainsi, comparativement à la méthode traditionnelle qui ne compte que les bactéries en mesure de croître sur un milieu de culture, la méthode à l'essai effectue une quantification du total des cellules bactériennes.

Il a été démontré dans la littérature que le compte des bactéries cultivables sous-estime grandement la population de bactéries vivantes [Bianchi et Giuliano, 1996; Newby, 2007]. De plus, le compte sur boîte de pétri sous-estime aussi le compte des bactéries cultivables dû à l'agrégation des cellules [Auty *et al.*, 2001; Reimann *et al.*, 2010; Salhani *et al.*, 1998]. Donc, il importe de connaître la proportion de bactéries comptées par la méthode traditionnelle parmi les bactéries totales dans les diverses substances pharmaceutiques afin de savoir comment interpréter les divergences de résultats entre la méthode traditionnelle et la méthode alternative.

Il importe de mentionner que les résultats de cette section seront rapportés en valeur relative. Le résultat de qPCR ou de compte sur pétri par gramme ou millilitre de produit sera divisé par la spécification du produit qui est elle aussi par gramme ou millilitre de produit. Ceci permet de respecter la confidentialité des résultats obtenus par la méthode traditionnelle et les valeurs des spécifications de l'usine sans nuire à l'interprétation des résultats.

9.2. Méthodologie

9.2.1. Essais sur l'eau purifiée

De l'eau provenant du circuit de purification d'eau de l'usine a été testée en parallèle par les deux méthodes. Deux contenants stériles étaient remplis à tour de rôle pour obtenir deux échantillons de 200 ml. Pour chaque méthode, 100 ml d'eau étaient utilisés. Trente et un échantillons ont été analysés en parallèle.

9.2.2. Essais sur les matières premières et produits finis

Le protocole a été adapté, pour faire l'analyse de matière première, tel que décrit dans la section 8.2. Six matières premières ont été analysées en duplicata selon l'adaptation du protocole pour l'analyse de ces matières tel que décrit au chapitre 8. Les échantillons de 100 grammes proviennent de matières premières du bâtiment de production, destinées aux analyses du laboratoire de microbiologies. Une fraction de cet échantillon est utilisée pour des tests conventionnels effectués par les analystes du laboratoire de microbiologie alors qu'une autre fraction est utilisée pour les tests de qPCR de cette étude. Les résultats de la méthode conventionnelle n'ont pas tous été obtenus. Les spécifications de l'usine de Pfizer Montréal pour ces matières serviront de références pour les matières dont le compte n'a pu être obtenu.

Trois produits finis ont été analysés en duplicata selon les adaptations du protocole détaillés au chapitre 8. Les échantillons utilisés proviennent de produits finis qui sont disponibles en pharmacie. Les résultats sur qPCR ont alors été comparés aux spécifications microbiologiques du produit.

9.2.3. Essais sur différents fournisseurs

Le carbonate de calcium de trois différents fournisseurs a été analysé. Premièrement, du CaCO_3 de grade laboratoire de chimie analytique (A.C.S) a été obtenu chez Fisher Scientific pour les fournisseurs Acros Organics (AC423511000, No.423515000) et MP Biomedicals (ICN19141380, No.0219141380). Le CaCO_3 a été commandé des fournisseurs en un seul exemplaire et a été testé en triplicata.

Deuxièmement, des échantillons de CaCO_3 de grade alimentaire d'un fournisseur approuvé ont été obtenus du laboratoire de microbiologie de l'usine. Les échantillons du fournisseur Huber ont été testés en parallèle par la méthode traditionnelle et par la méthode de qPCR. Deux variantes, en termes de numéro de produit, de CaCO_3 provenant de ce fournisseur ont été testées. La variante I a été testée pour trois lots différents et la variante II a été testée pour deux lots. Ainsi, cinq échantillons ont été testés pour ce fournisseur.

9.3. Résultats

9.3.1. Essais sur l'eau purifiée

Les moyennes pour les deux technologies sont en accord. Le résultat moyen des 31 tests d'eau est de 0,062 UCG relative le qPCR et 0,057 UFC pour le décompte traditionnel, Tableau 9.1.

Tableau 9.1. Sommaire des tests d'eau

Échantillons d'eaux n= 31	qPCR (UCG relative)	Décompte (UFC relative)
Moyenne	0.062	0.057
Écart-Type	0.053	0.050

*Les résultats du compte par PCR (UCG) et par la méthode traditionnelle (UFC) sont divisés par les spécifications de l'usine pour obtenir les unités relatives.

La concordance des résultats montre que pour l'eau purifiée, il n'y a pas de distinction entre le résultat obtenu avec la méthode traditionnelle et le résultat obtenu par qPCR. On peut supposer que le nombre de cellules mortes et le nombre de cellules viables non cultivables dans les tests d'eau peuvent être négligés. Ainsi le compte basé sur les bactéries cultivables seraient approximativement équivalents au compte basé sur l'ADN pour ce cas spécifique. La concordance de ces résultats suggère qu'il existe une équivalence entre les unités UFC-UCG, ce qui est en théorie difficile à obtenir étant donné les imperfections de ces deux unités discutées dans la section 4.1.3. Il est possible que cette équivalence soit due au fait que la méthode soit optimisée pour les tests d'eau. Ainsi, les standards d'ADN utilisés seraient fabriqués en vue d'obtenir une concordance entre l'UFC et l'UCG pour les tests d'eau. Il est aussi possible que l'impacte des différences entre les deux systèmes d'unités soit moins importants pour les tests d'eau purifiée. On retiendra que la divergence entre les bactéries cultivables et les bactéries totales, ne semble pas être applicable aux tests d'eau purifiés.

9.3.2. Essais sur les matières premières et produits finis

Les résultats exacts obtenus ne peuvent pas être diffusés pour des raisons de confidentialité, mais peuvent être examinés qualitativement. Les résultats de tests sur les matières premières (liquides et solides) ainsi que sur les produits finis montrent une différence marquée entre les résultats du qPCR et de la méthode traditionnels, ce qui suggère qu'un taux élevé d'ADN bactérien est trouvé dans les matières premières. Ces résultats sont non indicatifs de la qualité de la matière analysée : tous les échantillons ont passé les tests de qualité validés. La discussion, section 9.4, permettra de soulever les observations en provenance de ces résultats qui ne peuvent pas être divulgués.

9.3.3. Essais sur différents fournisseurs

Les résultats montrent que les échantillons de carbonate de calcium de grade réactif de laboratoire, Acros Organics et MP Biomedicals, ne semblent pas contenir de cellule en quantité significative. Le résultat de qPCR pour ces matières se trouve en dessous de la limite inférieure de quantification de 50 UGC par tests. Dans le cas du fournisseur de CaCO_3 de grade alimentaire, le résultat de quantification est très au dessus des échantillons de CaCO_3 de grade réactif. Le compte des UFC montre que tous les échantillons de CaCO_3 ont un compte qui respecte les spécifications.

9.4. Discussion

9.4.1. Analyse des matières à haute teneur en UCG

Pour toutes les substances sous forme solide testées ainsi que les matières premières liquides, le résultat de qPCR est significativement supérieur aux résultats basés sur la croissance sur pétri. Ainsi, la quantification par qPCR diverge significativement de l'énumération sur boîte de pétri. Ceci génère quatre questions :

1. Est-ce que la croissance est inhibée lors des tests de la méthode traditionnelle?
2. Est-ce que les tests de qPCR sont biaisés par la quantification directe de l'ADN extracellulaire (ADNe)?
3. Est-ce que les tests de qPCR sont biaisés par la quantification indirecte de l'ADNe?
4. Est-ce que le résultat de qPCR est biaisé par les CVNC et les cellules mortes?
5. Est-ce que le résultat de qPCR est biaisé par rapport au compte des colonies par les agrégats de cellules pouvant former des colonies, tel que discuté au chapitre 6 ?

Hypothèse d'inhibition de croissance

Les résultats des tests conventionnels pourraient être mis en doute en supposant que certaines composantes des matières pharmaceutiques inhibent la croissance des microorganismes. Cependant, cette hypothèse est rejetée puisque les analyses sur ces matières ont été effectuées par les analystes du laboratoire de microbiologie en utilisant des protocoles validés qui permettent de désactiver les substances pouvant inhiber la croissance. Les milieux de culture et les procédures sont testées de manière routinière pour leur capacité à promouvoir la croissance des microorganismes.

Hypothèse de la quantification directe de l'ADNe

Les résultats des tests de qPCR pourraient être mis en doute en supposant que la quantification est biaisée par l'ADN extracellulaire. La méthode, tel que décrite à la section 4.1.1 permet d'évacuer l'ADN extra cellulaire lors de l'étape de filtration des bactéries ce qui élimine la possibilité d'une quantification directe. Cependant il est possible que de l'ADNe provienne du liquide mouillant le filtre; ce liquide est originaire de l'échantillon. En effet, si la concentration en ADN extracellulaire de l'échantillon est très élevée, même si la majorité de l'ADN extracellulaire est évacué par le filtre, le liquide résiduel mouillant le filtre contiendra assez

d'ADN pour fausser les résultats. Cependant, cette hypothèse est invalide puisque 50 ml d'eau purifiés est poussé à travers le filtre afin d'évacuer l'ADN extracellulaire qui aurait pu subsister.

Hypothèse de la quantification indirecte de l'ADN_e

Des matières telles que le ginseng ou le carbonate de calcium peuvent encrasser légèrement le filtre, réduisant ainsi la taille des pores du filtre, ce qui pourrait permettre au filtre de retenir l'ADN_e. L'ADN serait alors récupéré lorsque le filtre subit l'étape de centrifugation puisque la pression exercée sur celui-ci est plus grande que lors de la filtration. Cette hypothèse pourrait expliquer les résultats du carbonate de calcium et du ginseng, mais pas celui du xylitol ou du sirop de maïs. Le xylitol est une petite molécule similaire au glucose et le sirop de maïs est un mélange de glucose et de fructose. Dans le cas du xylitol et du sirop de maïs, l'hypothèse de réduction de la taille des pores du filtre ne tient pas puisque le xylitol, le fructose et le glucose sont de petite molécule qui ne peut pas encrasser le filtre. Cette hypothèse est donc peu probable.

D'un autre côté il est possible que l'ADN extracellulaire soit présent en quantité suffisamment importante pour encrasser le filtre de façon durable de sorte que l'étape de rinçage avec l'eau purifiée ne soit pas suffisante pour évacuer cet ADN. Cette hypothèse plausible n'a pas été vérifiée pour des contraintes de temps et de matériel. Si tel est le cas, la quantification des cellules totales serait bruitée par la présence d'ADN extracellulaire.

Une autre possibilité serait que certaines molécules puissent s'adsorber sur le filtre sans nécessairement le colmater; ces molécules pourraient ensuite permettre l'adsorption de molécules d'ADN. Par la suite, la solution de lyse qui contient un surfactant pourrait déloger cet ADN du filtre qui serait par la suite quantifié.

Hypothèse du biais par les CVNC et les cellules mortes

Pour toutes les matières testées en parallèle avec la méthode traditionnelle, la quantification par qPCR est significativement supérieure au compte sur pétri. Ainsi, les matières premières et produits finis semblent contenir une population de CVNC, de cellules mortes et/ou d'ADN extracellulaire qui dépasse largement les cellules cultivables. Si l'hypothèse de l'ADN cellulaire quantifié de façon indirecte est fausse, alors le biais doit provenir des CVNC et des cellules mortes ou des agrégats cellulaires quantifiés comme un seul UFC. La littérature montre que les CVNC sont en général plus nombreuses que les cellules cultivables par plusieurs ordres de grandeur [Bianchi et Giuliano, 1996; Newby, 2007]. Ainsi, l'hypothèse d'un biais causé par les CVNC et cellules mortes est l'hypothèse la plus probable, sans toutefois être en mesure de rejeter l'hypothèse de l'encrassement du filtre par une grande concentration d'ADNe. Cette hypothèse pourrait être évaluée par une analyse en microscopie avec marquage de type *life/dead*.

9.4.2. Analyse des matières à basse teneur en UCG

De façon générale, les matières liquides analysées présentent moins d'ADN que les matières solides. On constate aussi que la charge d'UCG est moins grande pour les produits que pour les matières premières. Les charges d'UCG les moins élevées se trouvent dans l'eau purifiée; l'eau provenant du système de purification d'eau contient très peu d'UCG et la quantification est équivalente à celle de la méthode traditionnelle.

9.4.3. Analyse de différents fournisseurs

Le CaCO_3 de grade alimentaire contient 5 % de maltodextrine dans le but d'améliorer les propriétés mécaniques de la poudre; ceci explique peut-être la différence de résultats entre les grades alimentaires et analytiques. Il est possible que l'ajout de maltodextrine apporte quelques cellules bactériennes ou ADN extracellulaire puisque c'est un produit alimentaire; les produits alimentaires ne sont généralement pas stériles. Il est aussi possible que les procédés de purification du CaCO_3 , nécessaire à l'obtention de la pureté chimique du produit, aient pour effet d'éliminer les cellules bactériennes.

9.5. Sommaire

Les tests effectués sur des échantillons d'eau purifiée, de produits, de matières premières et de matières provenant de différents fournisseurs ont permis d'observer l'impact des cellules totales sur le résultat de qPCR et sur la divergence relativement aux tests de la méthode traditionnelle. Au niveau du résultat des tests, on peut classer les matières en quatre catégories en ordre croissante du résultats de qPCR :

- L'eau purifiée
- Les produits finis liquides
- Les matières premières liquides
- Les poudres

En résumé, la présence de bactéries viables non cultivables et de bactéries mortes n'affectent pas les résultats d'échantillons provenant du système de purification d'eau. On pourrait aussi supposer que les produits liquides sont peu susceptibles d'être affectés par ce phénomène. D'un autre côté les solides, matières premières et produits finis, sont probablement fortement affectés par ce phénomène.

Ainsi, les échantillons d'eau purifiée et certains produits liquides seraient des candidats intéressants pour l'utilisation du qPCR comme méthode alternative au test TAC. Pour pouvoir analyser les autres échantillons en utilisant la méthode, il faudrait investiguer davantage sur les liens qui unissent les bactéries cultivables, les CVNC, les bactéries mortes et l'ADN extracellulaire.

Une investigation approfondie des cellules totales permettrait d'établir si un contrôle est possible en suivant ceux-ci. En effet, il serait intéressant de savoir si la teneur en cellule totale est un aussi bon indicateur de salubrité, que le suivi des cellules cultivables. De plus, certains auteurs supposent que le suivi des bactéries viables non cultivables serait souhaitable, puisque celles-ci ont le potentiel de causer des pathologies [Newby, 2007]. Aussi, les bactéries mortes se trouvant dans les produits supposent que des métabolites bactériens s'y trouvent aussi.

10. CONCLUSION

10.1. Sommaire des travaux

L'analyse des performances analytiques de la méthode de quantification de bactérie par qPCR , chapitre 5, a montré que la méthode satisfait généralement aux critères de la United States Pharmacopeia. L'analyse s'est attardée aux critères de spécificité, d'exactitude, de linéarité, de précision, de robustesse et aux limites de quantification. Cependant, trois éléments devant être approfondis davantage ont été soulevés :

1. La relation entre les UCG et les UFC

La façon dont USP propose d'effectuer l'évaluation des critères analytiques de la méthode repose sur une comparaison entre les UCG et les UFC. Ce type de comparaison souffre du fait que le ratio UCG/UFC est fonction de l'espèce étudiée et de l'état du métabolisme cellulaire. Il est ainsi difficile de maintenir un ratio constant ce qui pose des défis pour l'interprétation des résultats. Des travaux supplémentaires ont été effectués dans le chapitre 6 pour investiguer l'impact du métabolisme cellulaire sur le rapport UCG/UFC.

2. La précision

Les résultats de précision sont à la limite de l'acceptabilité par rapport aux critères USP. Des travaux supplémentaires ont été effectués dans le chapitre 7 pour vérifier s'il est possible d'augmenter la précision en utilisant un algorithme de quantification supérieur.

3. L'effet des bactéries non cultivable

La présence de bactéries viables non cultivables, de bactéries mortes et d'ADN extracellulaire n'a pas été investiguée durant l'analyse des performances analytiques. Cependant, ce phénomène pourrait avoir un impact important sur les résultats avec des échantillons réels. Des travaux supplémentaires ont été effectués dans le chapitre 8 pour adapter la méthode et ainsi tester des échantillons réels pour évaluer l'impact des bactéries totales dans le chapitre 9.

L'analyse de l'effet de l'état métabolique sur le ratio UCG/UFC, chapitre 6, a permis d'apprendre que l'état métabolique des cellules dues à l'âge de la préparation bactérienne avait une influence directe sur le ratio UCG/UFC. L'hypothèse la plus probable est que le nombre de cellules par agrégats cellulaires diminue lorsque la phase de croissance des cellules atteint le stade de déclin ce qui affecte directement la quantification par PCR sans avoir un impact très important sur le compte des UFC. Ce phénomène pose des problèmes d'interprétation de résultats qui rendront la validation de la méthode difficile. Le problème peut être réglé de deux façons :

- Stabiliser le nombre de cellules par agrégats en ayant une préparation de cellules bactériennes qui donne des résultats reproductibles. Selon notre fournisseur, l'utilisation de *BioBall*[®] génère des résultats reproductible.
- Comparer les UCG à une méthode de quantification plus appropriée que les UFC. Le compte de cellules par microscopies avec marquage de type *live/dead* est un exemple.

L'effet de l'algorithme de quantification sur la précision de la méthode a été évalué au chapitre 7. Un algorithme supérieur au niveau de l'erreur, de la précision et du biais pour des variations inter-essais a été développé. L'algorithme a ensuite été substitué à l'algorithme du fournisseur pour l'analyse des données de fluorescence générée par la méthode de qPCR. L'algorithme a démontré une meilleure robustesse face aux problèmes d'inhibition de PCR. La précision mesurée pour les versions 1 et 2 du protocole s'est améliorée alors que la précision pour la dernière version du protocole est restée la même. On suppose que les essais effectués durant la dernière version du protocole utilisent des réactifs plus robustes. Aux finales, le nouvel algorithme n'a pas augmenté la précision de la méthode, selon le dernier protocole en vigueur, mais performera mieux lorsque soumis à des conditions difficiles. Le nouvel algorithme permet ainsi d'obtenir une méthode plus robuste. La précision devra donc être améliorée d'une façon différente.

- L'automatisation des étapes d'extraction et de purification d'ADN devrait permettre une augmentation de la précision.
- L'utilisation de répliques augmenterait la précision, accompagnée d'une augmentation du coût d'analyse par échantillon.

La méthode de qPCR a été évaluée pour sa capacité à être adaptée pour l'analyse de diverses matières pharmaceutiques (chapitre 8). Ces travaux ont aussi permis de préparer la méthode à l'analyse de diverses matières pharmaceutiques dans l'objectif d'évaluer le phénomène des bactéries viables non cultivables (chapitre 9). Sur les dix-huit matières investiguées, huit sont aisément adaptées pour les tests de qPCR. Les produits et matières contenant plusieurs matières insolubles en grande quantité dont la sédimentation n'est pas réalisable sont les matières les plus difficiles à adapter et demandent un développement de méthode spécifique.

- Dans une perspective d'implantation de la technologie, la méthode de qPCR devrait d'abord être considérée pour les tests d'eau purifiée, de liquide filtrable et les poudres solubles. Ensuite, la méthode pourrait faire l'objet d'un développement de méthode poussé pour l'analyse des autres types de matières.

Finalement l'effet des bactéries totales sur le résultat de la méthode a été investigué en comparant les résultats de tests de différent type de matière pour la méthode traditionnelle et la méthode de qPCR (chapitre 9). Il en ressort que les deux méthodes sont équivalentes pour les tests d'eau purifiée et sur les deux produits liquides testés. Cependant, la quantification en UCG pour les matières premières liquides et les matières solides est significativement différente de la méthode traditionnelle

- La méthode pourrait être validée pour les tests d'eau et certains produits liquides.
- La méthode pourrait être validée pour certains tests de matières premières liquides à condition de relever les spécifications.
- Une évaluation plus poussée sur le phénomène des bactéries totales, visant à démontrer que le suivi des bactéries totales permet un contrôle adéquat de la flore bactérienne, pourrait éventuellement permettre l'utilisation de la méthode pour les tests de matières solides en utilisant des spécifications en UCG.

10.2. Retour sur les objectifs et la question de recherche

10.2.1. Objectifs techniques

- I. Évaluer si le qPCR peut être utilisé pour remplacer les tests TAC**
 - Évaluer la méthode au niveau des normes pharmaceutiques USP et EP
 - Objectif réalisé au chapitre 5 : PERFORMANCES ANALYTIQUES
- II. Déterminer pour quel type de matière le qPCR peut remplacer les tests TAC**
 - Adapter la méthode pour l'analyse de différentes matières
 - Objectif réalisé au chapitre 8 : ADAPTABILITÉ DE LA MÉTHODE
 - Évaluer le principe d'équivalence comparativement à la méthode traditionnelle
 - Objectif réalisé au chapitre 9 : INFLUENCE DES BACTÉRIES TOTALES

10.2.2. Objectifs scientifiques

- I. Évaluer l'effet de l'état de viabilité des cellules sur la méthode**
 - Évaluer le biais créé par la présence de bactéries morte ou non cultivable dans les matières pharmaceutiques
 - Objectif réalisé au chapitre : 9 INFLUENCE DES BACTÉRIES TOTALES
 - Évaluer l'effet de l'état de croissance des cellules sur le biais entre les résultats de la méthode traditionnelle et les résultats de la méthode de qPCR
 - Objectif réalisé au chapitre : 6 EFFET DE L'ÉTAT MÉTABOLIQUE
- II. Évaluer l'effet de la méthode de quantification sur le résultat de quantification**
 - Objectif réalisé au chapitre 7 : EFFET DE L'ALGORITHME

10.2.3. Objectifs économiques

- I. Déterminer la rentabilité du projet**
 - Évaluer le coût des tests pour l'utilisation de la nouvelle méthode
 - Objectif non réalisé*
 - Se servir de l'usine de Pfizer Montréal comme cas de base et évaluer si l'implantation de l'appareil et de la méthode est justifiée économiquement
 - Objectif non réalisé*

*Les coûts et la rentabilité de la méthode de qPCR n'ont pas été évalués comme prévu puisque la technologie du fournisseur n'est pas aussi avancée commercialement que ce qui avait été prévu initialement. Le fournisseur doit automatiser les étapes d'extraction et de purification d'ADN ce qui fera augmenter le coût d'investissement, mais réduira les coûts d'utilisation (temps de main-d'œuvre). De plus, le coût final des réactifs et du matériel n'est pas ficelé. Puisque l'investissement en capital n'est pas connu et que le coût d'utilisation par tests ne l'est pas non plus, l'analyse économique ne peut être réalisée avec précision et serait par conséquent peu utile. Il a donc été décidé d'attendre que la technologie soit sur le marché, avant de calculer la rentabilité.

10.2.4. Question de recherche

Tel que discuté dans l'introduction, le projet de recherche doit répondre à la question suivante : est-ce que la méthode de qPCR proposée peut être utilisée pour remplacer les tests microbiologiques TAC en respectant l'aspect technique et économique?

Au niveau technique, la méthode de qPCR pourrait immédiatement être utilisée pour remplacer les tests TAC pour les tests d'eau et de certains produits liquides filtrables. Cependant, avant d'entreprendre les travaux de validation, une méthode de préparation de bactéries donnant des résultats UCG/UFC reproductible doit être validée. À l'intérieur d'une vision à long terme, il serait possible d'élargir les types de matières pouvant être utilisées en travaillant sur deux volets. Le premier volet consisterait à travailler sur le développement de méthode pour pouvoir extraire et purifier l'ADN d'une plus grande gamme de matière. Le second volet viserait à évaluer si le compte des bactéries totales permet d'effectuer un contrôle efficace de la flore microbiologique ce qui permettrait d'inclure les types de matières, dont la quantification l'ADN est différente de la quantification de UFC. Au niveau économique, la question n'a pas été évaluée pour les raisons mentionnées dans la section 10.2.3

10.3. Contributions originales

1. Les travaux présentés dans ce mémoire ont permis de faire l'évaluation d'une méthode de qPCR ayant l'objectif de remplacer le test du compte des bactéries totales pour le domaine pharmaceutique. L'évaluation est basée sur les lignes directrices de la United States Pharmacopeia en ce qui a trait à l'usage de méthode microbiologique alternative.
2. Les travaux ont aussi permis de soulever les lacunes en ce qui a trait à l'utilisation de la méthode traditionnelle comme base de comparaison. En effet, l'énumération des unités formant une colonie est biaisée par le phénomène d'agrégation. Ceci rend difficile la comparaison de la méthode traditionnelle à des méthodes capables de compter chaque cellule individuellement indifféremment de l'agrégation.
3. Les travaux présentés ont aussi permis de développer un algorithme de quantification d'ADN dont le principal avantage est la robustesse face aux variations inter-essais et aux phénomènes d'inhibition.

10.4. Nouvelles perspectives de recherches

1. Tel que discuté précédemment, la comparaison des UCG aux UFC pose certains problèmes. Il faudrait, soit comparer les UCG à une autre unité que les UFC, ou développer une méthode de préparation bactérienne qui permet d'obtenir un ratio UCG/UFC stable et répétable. Le compte par microscopie par exemple serait une meilleure base de comparaison que les UFC. L'utilisation de BioBall[®] est un exemple de préparation bactérienne permettant d'obtenir un ratio UCG-UFC repérable et stable selon notre fournisseur.
2. Tel que discuté précédemment, il serait intéressant d'évaluer si le compte de copies d'ADN totales permet d'effectuer un contrôle efficace de la flore microbiologique. Cela élargirait la portée d'utilisation du qPCR pour le type de matière testé.
3. Le RT-qPCR pourrait être développé pour la quantification des bactéries totales vivantes. L'avantage du RT-qPCR est qu'il est spécifique aux bactéries viables comparées au qPCR. Cependant, cette technologie possède plus de contraintes techniques.

ANNEXE A

Relation entre la pente de la courbe standard et l'efficacité idéale (E=1).

La pente de la courbe standard, A, est donnée par la formule suivante :

$$A = \frac{\Delta_{C_q}}{\Delta_{[DNA]}} = \frac{C_{q2} - C_{q1}}{\text{Log}\left(\frac{DNA_2}{DNA_1}\right)}$$

Lorsque l'efficacité est de 100 %, l'ADN double à chaque cycle, donc :

$$2DNA_2 = DNA_1 \text{ quand } E = 1$$

Donc quand pour deux échantillons dont le C_q varie de 1, l'équation de la pente devient :

$$A = \frac{1}{\text{Log}_{10}\left(\frac{1}{2}\right)} = -3,322$$

Ainsi, la courbe standard parfait possède une pente de -3,322.

ANNEXE B

Sommaire des changements au protocole

Protocole abrégé

1. Préparation de l'échantillon contenant des bactéries
2. Filtration de l'échantillon pour recueillir les bactéries
3. Ajout de solution de lyse
4. Sonication pour détruire les membranes cellulaire
5. Traitement thermique pour inactiver les DNAses
6. Centrifugation à travers un filtre pour recueillir l'ADN
7. Lavage de l'ADN avec divers solvant combiné à la filtration
8. Ajout de l'ADN sur plateau à PCR
9. Préparation du thermocycleur pour l'étape d'amplification d'ADN

Résumés des changements

1. Modification à l'origine du protocole #2
 - a. Changement de la procédure à l'étape 7. Modification du protocole de lavage de l'ADN tel que recommander par le fournisseur. Le fournisseur tient à ce que les détails restent confidentielles. La modification a permis d'éliminer des problèmes d'inhibition de PCR qui était présents lors de l'utilisation du protocole #1.
2. Modifications à l'origine du protocole #3
 - a. Changement de procédure à l'étape 8. Initialement, l'extrait d'ADN et le *master mix* étaient ajoutés au plateau de PCR sans être mélangés. Dans le protocole 3, une étape de mélange par pipetage répétitif a été ajouté.
 - b. Changement de la procédure à l'étape 9. Modification du nettoyage d'une pièce d'équipements réutilisable en contact avec l'ADN. Les changements ont permis de réduire le bruit de fond et la variabilité de celui-ci.
 - c. Changement de la procédure à l'étape 1. Modification au niveau de la préparation d'échantillons provenant de culture bactérienne. La nouvelle procédure permet une relation plus stable entre les UFC et les UCG.

11. RÉFÉRENCES

- Auty MAE, Gardiner GE, McBrearty SJ, O'Sullivan EO, Mulvihill DM, Collins JK, et al. Direct in situ viability assessment of bacteria in probiotic dairy products using viability staining in conjunction with confocal scanning laser microscopy. *Appl Environ Microbiol.* 2001;67(1):420-5.
- Bianchi A, Giuliano L. Enumeration of viable bacteria in the marine pelagic environment. *Appl Environ Microbiol.* 1996;62(1):174-7.
- Birch L, Dawson CE, Cornett JH, Keer JT. A comparison of nucleic acid amplification techniques for the assessment of bacterial viability. *Lett Appl Microbiol.* 2001;33(4):296-301.
- Boggy GJ, Woolf PJ. A mechanistic model of PCR for accurate quantification of quantitative PCR data. *PLoS ONE.* 2010;5(8).
- Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin Chem.* 2009;55(4):611-22.
- Carr AC, Moore SD. Robust quantification of polymerase chain reactions using global fitting. *PLoS ONE.* 2012;7(5).
- Coleman TF, Li Y. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization.* 1996;6(2):418-45.
- Dahllof I, Baillie H, Kjelleberg S. rpoB-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. *Appl Environ Microbiol.* 2000;66(8):3376-80.
- Denoya CD. Nucleic acid amplification - based rapid microbiological methods: Are these technologies ready for deployment in the pharmaceutical industry? *American Pharmaceutical Review.* 2009;12(4).
- Dorak T, Real-time PCR, Taylor & Francis Group, 2006.
- Dolan A, Burgess CM, Barry TB, Fanning S, Duffy G. A novel quantitative reverse-transcription PCR (qRT-PCR) for the enumeration of total bacteria, using meat micro-flora as a model. *J Microbiol Methods.* 2009;77(1):1-7.
- D'Urso OF, Poltronieri P, Marsigliante S, Storelli C, Hernández M, Rodríguez-Lázaro D. A filtration-based real-time PCR method for the quantitative detection of viable salmonella enterica and listeria monocytogenes in food samples. *Food Microbiol.* 2009;26(3):311-6.

Dusserre E, Ginevra C, Hallier-Soulier S, Vandenesch F, Festoc G, Etienne J, et al. A PCR-based method for monitoring legionella pneumophila in water samples detects viable but noncultivable legionellae that can recover their cultivability. *Appl Environ Microbiol*. 2008;74(15):4817-24.

Espy MJ, Uhl JR, Sloan LM, Buckwalter SP, Jones MF, Vetter EA, et al. Real-time PCR in clinical microbiology: Applications for routine laboratory testing. *Clinical Microbiology Reviews*. 2006;19(1):165.

Farajnia S, Hassan M, Hallaj Nezhadi S, Mohammadnejad L, Milani M, Lotfipour F. Determination of indicator bacteria in pharmaceutical samples by multiplex PCR. *Journal of Rapid Methods & Automation in Microbiology*. 2008;17(3):328.

Fey A, Eichler S, Flavier S, Christen R, Höfle M, Guzman C. Establishment of a Real-Time PCR-Based Approach for Accurate Quantification of Bacterial RNA Targets in Water, Using *Salmonella* as a Model Organism. *Appl. Environ. Microbiol*. 2004; 70(6):3618

Franke M, Thoben K-, Söller R. Approach to enable an automatic pre-processing of qRT-PCR: Analysis. ; 2012.

Goll R, Olsen T, Cui G, Florholmen J. Evaluation of absolute quantitation by nonlinear regression in probe-based real-time PCR. *BMC Bioinformatics*. 2006;7.

Guescini M, Sisti D, Rocchi MBL, Stocchi L, Stocchi V. A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition. *BMC Bioinformatics*. 2008;9.

Guescini M, Sisti D, Rocchi MBL, Panebianco R, Tibollo P, Stocchi V. Accurate and precise DNA quantification in the presence of different amplification efficiencies using an improved Cy0 method. *PLoS ONE*. 2013;8(7).

Haugland RA, Siefring SC, Wymer LJ, Brenner KP, Dufour AP. Comparison of enterococcus measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis. *Water Res*. 2005;39(4):559-68.

Hierro N, Esteve-Zarzoso B, Mas A, Guillamón JM. Monitoring of *saccharomyces* and *hanseniaspora* populations during alcoholic fermentation by real-time quantitative PCR. *FEMS Yeast Research*. 2007;7(8):1340-9.

Higuchi R, Dollinger G, Walsh PS, Griffith R. Simultaneous amplification and detection of specific DNA sequences. *Bio/Technology*. 1992;10(4):413-7.

Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR analysis: Real-time monitoring of DNA amplification reactions. *Bio/Technology*. 1993;11(9):1026-30.

Huggett J, Bustin SA. Standardisation and reporting for nucleic acid quantification. *Accreditation and Quality Assurance*. 2011;16(8):399-405.

Jimenez L, Smalls S, Ignar R. Use of PCR analysis for detecting low levels of bacteria and mold contamination in pharmaceutical samples. *J Microbiol Methods*. 2000;41(3):259-65.

Jimenez L. Molecular applications to pharmaceutical processes and cleanroom environments. *PDA journal of pharmaceutical science and technology*. 2011;65(3):242.

Karlen Y, McNair A, Perseguers S, Mazza C, Mermod N. Statistical significance of quantitative PCR. *BMC Bioinformatics*. 2007;8.

Kephart CM, Bushon RN. Evaluation of real-time quantitative polymerase chain reaction (qPCR) to determine escherichia coli concentrations at two lake erie beaches. *Scientific Investigations Reports*. 2009.

Life Technologies Corporation (2012), Real-time PCR handbook

Liu W, Saint DA. Validation of a quantitative method for real time PCR kinetics. *Biochem Biophys Res Commun*. 2002;294(2):347-53.

Martinon A, Cronin UP, Quealy J, Stapleton A, Wilkinson MG. Swab sample preparation and viable real-time PCR methodologies for the recovery of escherichia coli, staphylococcus aureus or listeria monocytogenes from artificially contaminated food processing surfaces. *Food Control*. 2012;24(1-2):86-94.

Monteiro L, Bonnemaïson D, Vekris A, Petry KG, Bonnet J, Vidal R, et al. Complex polysaccharides as PCR inhibitors in feces: Helicobacter pylori model. *J Clin Microbiol*. 1997;35(4):995-8.

Mullis K, Nobel lecture: The polymerase chain reaction, Nobelprize.org, 1993, http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1993/mullis-lecture.html.

Nadkarni MA, Martin FE, Jacques NA, Hunter N. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology*. 2002;148(1):257-66.

Newby P. The significance and detection of VBNC microorganisms. *American Pharmaceutical Review*. 2007;10(4).

Osborne JW. Improving your data transformations: Applying the box-cox transformation. *Practical Assessment, Research & Evaluation*. 2010;15(12):1.

Pathak S, Awuh JA, Leversen NA, Flo TH, Åsjø B. Counting mycobacteria in infected human cells and mouse tissue: A comparison between qPCR and CFU. *PLoS ONE*. 2012;7(4).

Paradis S, Boissinot M, Paquette N, Bélanger SD, Martel EA, Boudreau DK, et al. Phylogeny of the enterobacteriaceae based on genes encoding elongation factor tu and F-ATPase β -subunit. *Int J Syst Evol Microbiol*. 2005;55(5):2013-25.

Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 2001;29(9).

Postollec F, Falentin H, Pavan S, Combrisson J, Sohier D. Recent advances in quantitative PCR (qPCR) applications in food microbiology. *Food Microbiol.* 2011;28(5):848-61.

Reimann S, Grattepanche F, Rezzonico E, Lacroix C. Development of a real-time RT-PCR method for enumeration of viable bifidobacterium longum cells in different morphologies. *Food Microbiol.* 2010;27(2):236-42.

Ruijter JM, Pfaffl MW, Zhao S, Spiess AN, Boggy G, Blom J, et al. Evaluation of qPCR curve analysis methods for reliable biomarker discovery: Bias, resolution, precision, and implications. *Methods.* 2013;59(1):32-46.

Rutledge RG, Côté C. Mathematics of quantitative kinetic PCR and the application of standard curves. *Nucleic Acids Research.* 2003;31(16).

Rutledge RG. Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. *Nucleic Acids Res.* 2004;32(22).

Salhani N, Uelker-Deffur A. Improved quantification of aggregated bacteria by combined enzymatic and mechanical treatment of flocs and biofilm from a rotating drum bioreactor. *Water Res.* 1998;32(4):1287-95.

Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem.* 1964;36(8):1627-39.

Schafer RW. What is a savitzky-golay filter? *IEEE Signal Process Mag.* 2011;28(4):111-7.

Schrader C, Schielke A, Ellerbroek L, Johne R. PCR inhibitors - occurrence, properties and removal. *J Appl Microbiol.* 2012;113(5):1014-26.

Sisti D, Guescini M, Rocchi MBL, Tibollo P, D'Atri M, Stocchi V. Shape based kinetic outlier detection in real-time PCR. *BMC Bioinformatics.* 2010;11.

Sivaganansan M, Varma M, Haugland RA. Comparison of enterococcus quantitative polymerase chain reaction analysis results from fresh and marine waters on two real-time instruments. *Anal Biochem.* 2012;430(1):68-74.

Smith C, Osborn, A. Advantages and limitations of quantitative PCR(Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol Ecol* 67 (2009) 6–20

Spiess A-, Feig C, Ritz C. Highly accurate sigmoidal fitting of real-time PCR data by introducing a parameter for asymmetry. *BMC Bioinformatics.* 2008;9.

Sutton S. The Limitations of CFU : Compliance to CGMP Requires Good Science, Journal of GXP Compliance. 2012;16(1)

Tanaka Y, Takahashi H, Simidu U, Kimura BON. Design of a new universal real-time PCR system targeting the *tuf* gene for the enumeration of bacterial counts in food. J Food Prot. 2010;73(4):670-9.

Tichopad A, Dilger M, Schwarz G, Pfaffl MW. Standardized determination of real-time PCR efficiency from a single reaction set-up. Nucleic Acids Res. 2003;31(20).

Trampuz A, Piper KE, Steckelberg JM, Patel R. Effect of gamma irradiation on viability and DNA of staphylococcus epidermidis and escherichia coli. J Med Microbiol. 2006;55(9):1271-5.

United states pharmacopeia and national formulary. Rockville, MD, USA: United States Pharmacopeia Convention; 2012.

Walker MJ, Montemagno C, Bryant JC, Ghiorse WC. Method detection limits of PCR and immunofluorescence assay for cryptosporidium parvum in soil. Appl Environ Microbiol. 1998;64(6):2281-3.

Wang S, Li Q-, Li Z, Su M-. Solubility of xylitol in ethanol, acetone, N,N-dimethylformamide, 1-butanol, 1-pentanol, toluene, 2-propanol, and water. J Chem Eng Data. 2007;52(1):186-8.

Wilhelm, J et Pingoud, A. Real-time polymerase chain reaction. ChemBioChem. 2003;4(11):1120-8.

Wilhelm J, Pingoud A, Hahn M. (2003a), Validation of an algorithm for automatic quantification of nucleic acid copy numbers by real-time polymerase chain reaction. Analytical Biochemistry. 2003;317(2):218-25.

Wilhelm J, Pingoud A, Hahn M. (2003b), SoFAR: Software for fully automatic evaluation of real-time PCR data. BioTechniques. 2003;34(2):324-32.

Yamamoto S, Harayama S. PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of pseudomonas putida strains. Appl Environ Microbiol. 1995;61(3):1104-9.

